

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta
Ústav aplikované informatiky

Dynamika sociálních sítí

Diplomová práce

Bc. Roman Klimeš

Školitel: Ing. Jiří Jelínek, CSc.

České Budějovice 2015

Bibliografické údaje

Klimesš, Roman, 2015: Dynamika sociálních sítí. [Social Networks Dynamics. Mgr. Thesis, in Czech.] – 76 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Anotace

Sociální sítě patří v současné době díky své popularitě mezi uživateli k nejvíce studovaným strukturám v oblasti výměny informací a znalostí. V této práci se budeme věnovat dynamice šíření informací a znalostí v těchto sítích. Bude představen zdokonalený heuristický model chování jedince v síti, založený na jeho potřebě získávání informací a znalostí pro řešení konkrétních situací. Z něho vychází agent-based model celé sociální sítě, který byl prakticky implementován, a prezentovány budou rovněž jeho výstupy při simulaci vybraných úloh testujících dynamiku sociální sítě.

Abstract

Social networks are currently the most studied structures due to its popularity among users. In this thesis we will focus on the dynamics of the dissemination of information in these networks. There will be introduced the advanced heuristic model of individual behavior in the network which is based on individuals need for information and knowledge for solving specific situations in this paper. The multi-agent model of the social network, which was adapted and expanded and also practically implemented, is based on this conceptual model, and its outputs from the simulations of selected tasks testing the dynamics of social networks will be presented as well.

Prohlášení

Prohlašuji, že svoji diplomovou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 14. 12. 2015

Podpis

Poděkování

Děkuji panu Ing. Jiřímu Jelínkovi, CSc., za odborné vedení mé diplomové práce a pomoc při jejím zpracování.

Obsah

1 Úvod	6
2 Komplexní sítě	7
2.1 Sociální síť	7
2.2 Vlastnosti a metriky komplexních sítí	10
2.3 Modely komplexních sítí	16
3 Dynamika sociálních sítí	21
3.1 Pojem dynamiky a její význam	21
3.2 Aktuální studie v dynamice sociálních sítí	22
3.3 Metodiky modelování a zkoumání dynamiky	27
4 Použitý model	32
4.1 Popis modelu	32
5 Úpravy modelu	34
5.1 Použitý software a transformace do nové verze	34
5.2 Model jako komplexní síť	36
5.3 Chování agenta	38
5.4 Eliminace předpokladu uzavřeného světa	41
6 Výsledky experimentů	45
6.1 Rozhodování agenta	45
6.2 Úprava ochoty odpovídat a snižování ratingu	47
6.3 Partnerství jako efekt zodpovězení dotazu	49
6.4 Rostoucí síť a upřednostňované připojování	53
6.5 Flexibilní počet situací	59
6.6 Analýza metodikou časových řad	62
7 Závěr	70
Seznam použité literatury	72
Seznam obrázků	74
Seznam tabulek	75
Příloha	76

1 Úvod

Sociální sítě patří v současné době k nejvíce studovaným strukturám v oblasti výměny informací a znalostí, přičemž častěji je zkoumáno jejich statické chování, poněkud méně pak jejich dynamika.

Sociální sítí rozumíme jakoukoli propojenou skupinu lidí, ve které jsou osoby propojeny vzájemnými vazbami. Jde tedy o graf vzájemně propojených uzlů, kde uzly představují jedince v síti a hrany grafu vztah mezi nimi. Dynamika sítě je pak dána zejména změnami v počtu uzlů a vazeb a změnách chování uzlů či umístění vazeb v čase.

Sociální sítě samotné nejsou produktem IT, ale jsou těmito technologiemi více či méně podporovány. Příkladem jednodušší verze podpory mohou být např. komunikační nástroje jak pro offline, tak online komunikaci. Vyšší úroveň pak představují celá prostředí umožňující mimo komunikaci i správu osobní sociální sítě a její rozšiřování. V tomto případě se již dá hovořit o online sociálních sítích [1]. Typickými systémy pro komplexní podporu online sociálních sítí jsou např. Facebook, MySpace a řada dalších.

Tato diplomová práce si klade za cíl rozvinout a zdokonalit již dříve navržený model pro simulaci dynamiky šíření informací a znalostí v sociální síti a na experimentálních výsledcích demonstrovat vliv provedených změn i některé další jevy, které se v dynamice sítí projevují.

Před samotnými modifikacemi modelu se bude práce věnovat pojmu komplexní sítě, pod který sociální sítě spadají, jejím základním vlastnostem, metrikám a analýze, a pozornost bude věnována také teoretickým modelům síťové struktury. V další části práce se budeme zabývat aktuálními informačními zdroji zaměřenými na dynamiku sociálních sítí a zaměříme se rovněž na metodiky modelování a zkoumání dynamiky.

S přihlédnutím k nashromážděným teoretickým poznatkům se poté, jak již bylo zmíněno, pokusíme rozvinout zvolený model, který bude nejprve podrobně popsán, stejně tak jako všechny jeho následné úpravy. Na konci práce bude potom provedena řada experimentů, kde bude s využitím popsaných metodik snahou ukázat vliv provedených modifikací na dynamiku vývoje sítě.

2 Komplexní síť

S rozvojem informačních technologií, kdy jsme schopni mimo sítě o několika desítkách, maximálně stovkách uzlů, jež byly sestaveny na základě dotazníků, pracovat i se sítěmi o milionech a mnohdy i miliardách uzlů, se dostáváme k pojmu komplexní síť.

Komplexní síť je, jak již označení napovídá, složitá síť s rozsáhlým grafem. Podle [2] můžeme komplexní síť shrnout do následujících čtyř základních kategorií, z nichž nás bude nejvíce zajímat ta první, a tou jsou sociální síť.

- **Sociální síť** – síť lidí udržující nějaký vzájemný vztah, např. Facebook
- **Informační síť** – např. Web nebo síť citací mezi akademickými publikacemi
- **Technologické síť** – např. síť pro distribuci elektřiny, tepla atd. a také internet
- **Biologické síť** – vše, co lze aplikovat na biologický systém, např. neuronová síť

2.1 Sociální síť

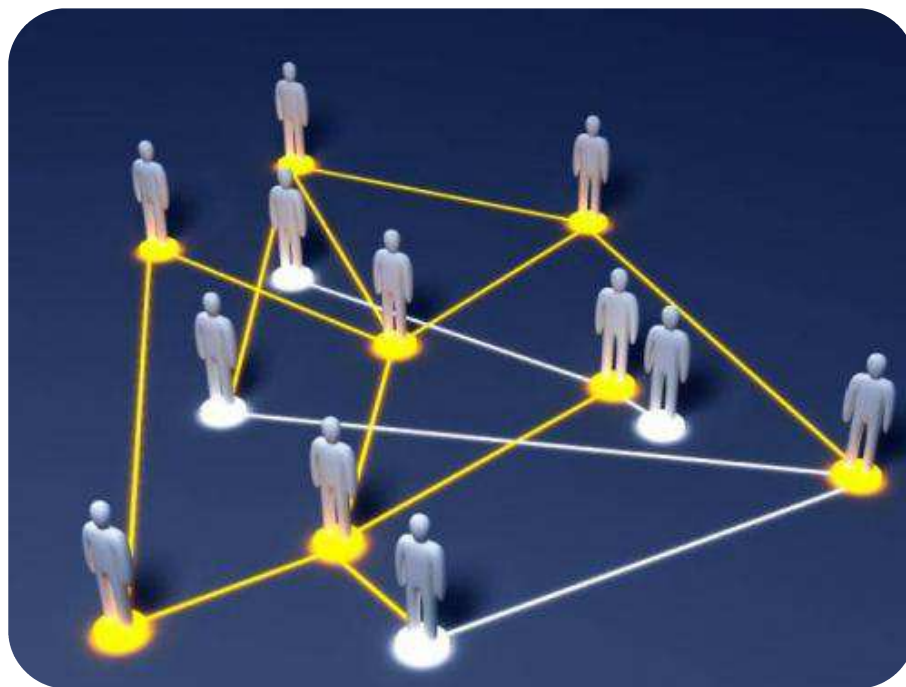
Sociální síť je pojem dnes již hojně využívaný především v informatice a v prostředí internetu. Takovéto chápání ovšem není zcela výstižné a může být poněkud zavádějící. Je třeba si uvědomit, že tento pojem existoval již dlouho před zrozením internetu a vůbec informatiky jako takové. Jako první ho použil James A. Barnes v oblasti sociologie, který se zajímal o malé skupiny a popis jejich vztahů.

Sociální síť rozumíme jakoukoli propojenou skupinu lidí, kterou dohromady spojuje určitá společná skutečnost. Může to být příbuzenský vztah, práce nebo třeba společný koníček. Nejtypičtějším příkladem takové sociální sítě je samozřejmě rodina. Je to první sociální síť, ve které se člověk nachází hned po narození. Sociální síť je možné si představit jako graf vzájemně pospojovaných uzlů, kde jednotlivé uzly představují jedince a propojení uzlů vztah mezi nimi (Obr. 1).

2.1.1 Základní vlastnosti a analýza sociálních sítí

Jak uvádí [3], analýza sociálních sítí je způsob šetření, který se zaměřuje na vztahy mezi subjekty, jako jsou jednotlivci nebo organizace. Nicméně je využívána i ke zkoumání vztahů mezi objekty tak rozmanitými jako je internet, vědecké práce, organismy a molekuly. Sociální síť je předmětem rostoucího zájmu vědy, stejně tak jako sociální vědy v posledních letech.

Podle [1] můžeme sociální síť rozdělit do dvou odlišných kategorií, a to na *offline* a *online* sociální síť. Termínem „offline“ myslíme všechny sociální síť s tradiční, tzn. nedigitální komunikací. Na druhou stranu „online“ odkazuje na sociální síť s jakýmkoliv typem digitální komunikace (například e-mail, webové aplikace, telefonní hovory).



Obr. 1: Sociální síť jako graf spojených uzlů

Zdroj: http://www.inflow.cz/files/redakce/Socialni_site.jpg

Z hlediska topologie sítě rozlišujeme síť (grafy) symetrické neboli neorientované, kde vazba mezi uzly nemá žádný směr, to znamená, že vazba (a,b) je stejná jako vazba (b,a), a síť orientované, kde má každá vazba jeden ze dvou směrů. V tom případě jsou vazby (a,b) a (b,a) dvě různé vazby.

Zmiňme ještě to, že ne všechny vazby v síti mají stejnou váhu. Potom mluvíme o síti s váženými vazbami, kdy má každá vazba kromě případného orientovaného směru ještě váhu (hodnotu), která určuje její sílu. Definicí síly vazeb se hodně zabýval americký sociolog Mark Granovetter. [4]

Lze usoudit, že orientovaný typ sítě s hodnocenými vazbami je zobecněným typem pro odlišné typy sítí, protože neorientovaná síť může být reprezentovaná sítí orientovanou, kde původní neorientovanou vazbu představují dvě vazby orientované, a nehodnocené vazby mohou být jednoduše reprezentovány vazbami hodnocenými, přičemž váhy všech vazeb budou totožné a budou se rovnat 1.

Důležitou vlastností sociální sítě je průměrná vzdálenost neboli průměrná nejkratší cesta mezi jakýmkoliv uzly v síti, která bude podrobněji popsána později. Při hledání nejkratší cesty v síti téměř jistě narazíme na *Dijkstrův algoritmus* [5], který je pro toto hodně používaný a pomocí něhož vytvoříme strom nejkratších cest z jednoho uzlu ke všem ostatním uzlům. Zajímavou vlastností nejkratší cesty je to, že jakákoliv její část je také nejkratší cestou.

S průměrnou délkou cesty v síti souvisí i fenomén tzv. „malého světa“. Každý z nás jistě zná větu „*Ten svět je ale malý.*“ v případě, že potkáme někoho známého na místě, kde bychom ho nečekali anebo někde daleko od domova potkáme někoho a po chvíli zjistíme, že máme společné známé. Přesně taková „malost“ světa je myšlena tímto fenoménem. Průměrné nejkratší cesty mezi uzly v komplexních sítích jsou jednoduše malé.

Jak uvádí autoři v [1], Stanley Milgram ve svém slavném experimentu v roce 1967 zjistil, že typická vzdálenost v sociální síti je kolem 6. Tato vlastnost je lépe známá právě jako „malý svět“ (small world) nebo „šest stupňů oddělení“ (six degrees of separation). Dokonce byl v roce 1993 natočen i stejnojmenný film [6].

Experiment spočíval v tom, že Milgram náhodně vybral několik desítek lidí v USA a požádal je, aby předali balíček nebo dopis osobě jim vzdálené, ovšem pouze v případě, že si navzájem tykají. V opačném případě se měli pokusit najít mezi svými známými (známost opět na úrovni tykání) tu nejbližší adresátovi a požádat ji o totéž. K překvapení všech z experimentu vyplynulo, že každá dvojice lidí v USA je spojena řetězem spojujícím průměrně 6 osob. Experiment samozřejmě obsahoval spoustu nedostatků a nemusí být zcela přesný, i tak lze ovšem předpokládat, že průměrné vzdálenosti v sociálních sítích, které dnes zahrnují i miliardy uzlů, budou velmi malé.

S dosažitelností uzlů neboli s tím, zda existuje alespoň jedna cesta mezi všemi páry uzlů, souvisí propojenost sítě. Z tohoto hlediska můžeme síť (grafy) rozdělit na propojené a nepropojené. Nepropojené části sítě se potom nazývají komponenty. Propojená síť neboli síť, kde jsou všechny páry uzlů dosažitelné, se logicky skládá pouze z jedné komponenty, která zahrnuje celou síť.

Základní metrika nebo minimálně metrika, které se při analyzování sociálních sítí určitě nevyhneme, je důležitost uzlu v síti neboli *centralita uzlu*. Způsobů, jak tuto vlastnost měřit, existují k dnešnímu dni stovky [7], avšak nejpoužívanější z nich jsou *degree centrality* (centralita měřená stupněm uzlu), *closeness centrality* (centralita měřená blízkostí polohy ve středu) a *betweenness centrality* (centralita měřená středovou mezípolohou). Podrobnější popis těchto tří centralit můžeme nalézt v [3].

Vlastností sociálních sítí bychom mohli nalézt mnohem více. Každá sociální síť má svou velikost (počet uzlů), hustotu propojení (počet vazeb), robustnost (počet možných cest) a mnoho dalších. Můžeme ještě zmínit to, že všechny uzly v síti nemusí spadat pouze do jedné kategorie, ale lze je rozdělit do více kategorií, např. servery, routery, PC nebo v případě sociálních sítí potom např. muži a ženy nebo děti, dospělí a důchodci apod. Takové sítě se nazývají *multimódové*. Stejně jako v případě uzlů i vazby v síti mohou mít více typů jako např. vazba znázorňující přátelství, nepřátelství, známost, kolegiální nebo příbuznost. Síť s více typy vazeb nazýváme *multiplexní*.

Studium sociálních sítí začalo analýzou offline sítí, typicky studiem dat z dotazníků a rozhovorů. Takovéto sociální sítě byly primárně analyzovány v sociologii, antropologii a evoluční psychologii. [1]

Více o analýze sociálních sítí, jejich metodách, softwarových nástrojích, s jejichž pomocí ji lze provádět, a příklad takové analýzy můžeme nalézt v bakalářské práci zde [3]. Velice rozsáhlá a detailní, i když už ne tak aktuální, ovšem jistě stále velmi použitelná, publikace zabývající se analýzou sociálních sítí je zde [8].

2.2 Vlastnosti a metriky komplexních sítí

Komplexní sítě mají typické vlastnosti jako je shlukování, distribuce stupňů, malé vzdálenosti atd. Při zkoumání dynamiky sociálních sítí se můžeme setkat s celou řadou takových metrik. Pojďme si je uvést.

2.2.1 Koeficient shlukování (Clustering coefficient)

Velice často měřená vlastnost, s níž se lze setkat u mnoha studií zabývajících se sociálními sítěmi, např. [1], [7], [9] a [10].

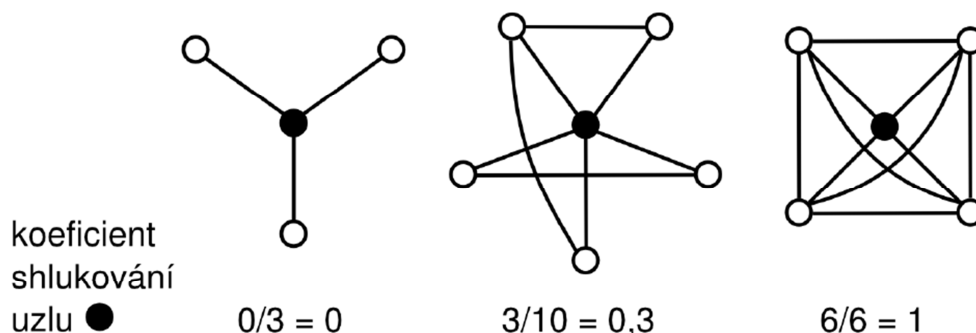
Koeficient shlukování udává to, zda nebo jak moc jsou sousedé nějakého uzlu propojeni neboli jednoduše to, zda se vaši přátelé navzájem znají. Lidé mají tendenci tvořit shluky, to znamená, znám-li Pepu a Frantu, je pravděpodobné, že Pepa bude znát Frantu.

Pojďme se na shlukování podívat formálněji:

- uzel v má k_v sousedů
- e_v je počet vzájemně propojených dvojic sousedů
- koeficient shlukování je potom definován vzorcem (1)

$$Cv = \frac{e_v}{\binom{k_v}{2}} = \frac{2e_v}{k_v(k_v - 1)} \quad (1)$$

Je to tedy poměr počtu dvojic sousedů, kteří jsou propojeni, a celkového možného počtu dvojic sousedů. Názornou ukázkou je možné zhlédnout na Obr. 2. Koeficient shlukování je měřen vždy pro černý uzel.



Obr. 2: Ukázka výpočtu koeficientu shlukování

Zdroj: [11]

Koeficient shlukování nabývá hodnot z intervalu $\langle 0;1 \rangle$, přičemž hodnoty 0 nabyde v případě, že ani jedna dvojice sousedů není propojena (první případ v Obr. 2) a hodnoty 1 v případě, že jsou propojeny všechny dvojice sousedů (poslední třetí případ v Obr. 2).

2.2.2 Distribuce stupňů (Degree distribution)

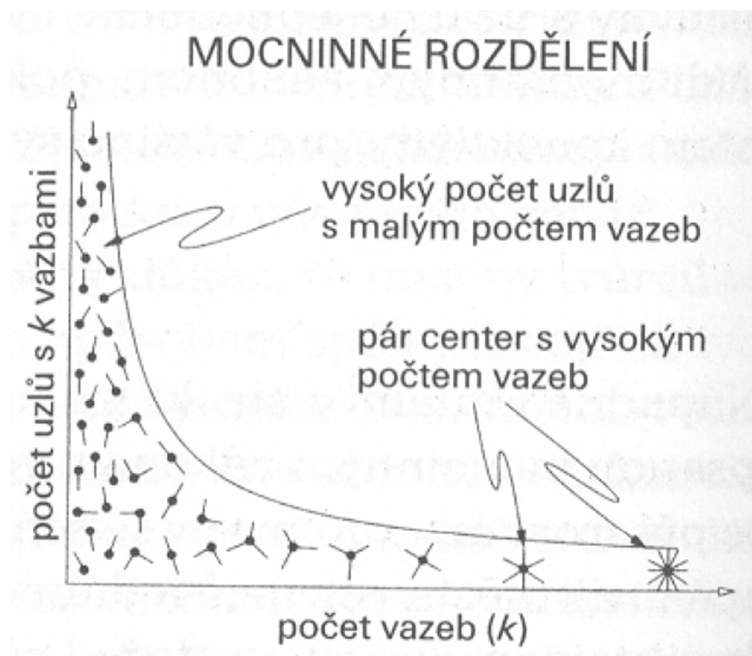
S distribucí stupňů se můžeme setkat také velice často, např. v [1] a [7]. Formálně je distribuce stupňů pravděpodobnost $P(k)$, že náhodně vybraný uzel má stupeň k .

Položme si otázku, zda existuje typická hodnota, kolem které se stupeň uzlů pohybuje, tzv. „škála grafu“. U náhodných grafů, které budou popsány dále, ano, ovšem u reálných komplexních sítí tomu tak není. Rozdělení stupňů v komplexních sítích lze velice často popsat tzv. mocninným rozdělením neboli mocninným zákonem, jehož znázornění je možné vidět na Obr. 3. Mocninný zákon je popsán vzorcem (2)

$$P(k) \sim k^{-\gamma} \quad (2)$$

což znamená, že většina uzlů v síti má podobně nízký stupeň a malá skupina uzlů naopak výrazně vysoký stupeň. Podíváme-li se pozorně na vzorec mocninného zákona, zjistíme, že

se jedná o lomenou funkci (základní $\frac{1}{x}$, v našem případě $\frac{1}{k^\gamma}$), jejímž grafem je hyperbola. Výsledný tvar hyperboly závisí na parametru γ , který je větší než 1 a obvykle v rozmezí $2 < \gamma < 3$.



Obr. 3: Distribuce stupňů podle mocninného zákona (Power law)

Zdroj: [11]

Aniž o tom víme, na mocninný zákon můžeme narazit i v běžném životě, například u výskytu křestních jmen v populaci. Existuje několik málo velmi frekventovaných jmen (Josef, Jan, Václav atd.). Zbýlá jména, a že jich je dost, se vyskytují výrazně méně. Dalším příkladem může být velikost měst. Většina měst v České republice má malou a srovnatelnou velikost, ale taková města jako je Praha, Brno nebo Ostrava spočítáme pomocí deseti prstů. Stejně tak je to s rozdělením bohatství. Většina lidí na světě je velmi chudá a několik málo jich je extrémně bohatých.

V komplexních sítích je tedy distribuce stupňů taková, že máme několik uzlů s výrazně vysokým stupněm (hodně propojené) a většinu ostatních s nízkým stupněm (málo propojené). Příklad mohou být webové stránky nebo sítě citací. Sítím, které sledují mocninný zákon, se říká *scale-free* (bezškálové), protože u nich neexistuje typický stupeň uzlu.

2.2.3 Průměrná vzdálenost (Average path length)

Průměrná vzdálenost je další často používaná metrika. Lze se s ní setkat např. v [7] a [9]. Jde o průměrnou vzdálenost nebo, chceme-li, průměrný počet skoků mezi všemi páry uzlů v síti. Jednoduše jde o aritmetický průměr všech nejkratších cest neboli o poměr součtu všech nejkratších cest a počtu těchto cest. Průměrnou vzdálenost lze získat podle vzorce (3)

$$l_G = \frac{1}{n \cdot (n - 1)} \cdot \sum_{i \neq j} d(v_i, v_j) \quad (3)$$

kde n je počet uzlů grafu G a $d(v_i, v_j)$ je nejkratší vzdálenost mezi uzly v_i a v_j . Průměrná vzdálenost může být použita například jako míra efektivity pro informační šíření v síti nebo jako míra efektivity hromadné dopravy.

2.2.4 Výstřednost (Eccentricity)

Se vzdáleností souvisí i výstřednost. Výstřednost $\epsilon(v)$ uzlu v je největší vzdálenost mezi uzlem v a jakýmkoliv jiným uzlem neboli výstřednost $\epsilon(v)$ určuje, jak daleko je uzel v od nejvzdálenějšího uzlu v grafu. Čím nižší je výstřednost, tím blíže středu grafu se uzel nachází (je vystředěnější).

2.2.5 Poloměr sítě (Radius)

Poloměr r grafu je minimální výstřednost mezi všemi uzly, formálně definován vzorcem (4)

$$r = \min_{v \in V} \epsilon(v) \quad (4)$$

to znamená, je to vzdálenost určená nejvzdálenějším uzlem od nejvystředěnějšího uzlu v grafu.

2.2.6 Průměr sítě (Diameter)

Další, co lze u sítí měřit, je jejich průměr, který taktéž souvisí se vzdáleností. Průměr d grafu je naopak maximální výstřednost mezi všemi uzly, formálně definován vzorcem (5)

$$d = \max_{v \in V} \epsilon(v) \quad (5)$$

to znamená, není to nic jiného než nejdelší cesta z nejkratších cest mezi všemi dvojicemi uzlů. Hodnota průměru grafu by se logicky měla rovnat zhruba dvojnásobku poloměru grafu.

2.2.7 Homofilní rysy (Assortativity)

Homophily („love of the same“, tj. „láska k těm, kdo jsou stejní“) je tendence individuí navazovat vazby s těmi, kteří jsou nějakým způsobem podobní. To může zahrnovat například věk, pohlaví nebo profesi. V oblasti sociálních sítí je tato metrika známá jako *assortativity* nebo *assortative mixing*.

Jak je uvedeno v [10], sociální sítě jsou typicky *assortative*, což znamená, že vzájemný vztah stupňů (degree) uzlů je pozitivní, to znamená, uzly s vysokým stupněm mají tendenci navazovat vazby s jinými uzly s vysokým stupněm a naopak. Toto zjištění bylo podpořeno několika studiemi o speciální struktuře a chování sociálních sítí, které je odlišily od jiných nesociálních sítí.

Druhou zvláštní charakteristikou, která byla studovaná a přispěla k vysvětlení *assortativity* sociálních sítí, je podle [10] shlukování (clustering), které již bylo popsáno výše. Shlukování bylo definováno z hlediska tranzitivity sítě, tzn., máme-li vazbu mezi párem uzlů A a B a jinou vazbu mezi uzly A a C, řekneme, že síť je silně tranzitivní v případě, že je pravděpodobné, že bude vazba i mezi párem uzlů B a C. Bylo vypořezováno, že koeficient shlukování u sociálních sítí je typicky větší než u odpovídajících náhodných grafů. Tento efekt byl dlouho studovaný sociology, kteří ho nazývali „*triadic closure bias*“, tzn. tendence jedinců navazovat vazby s přítelem přítele raději, než udržovat vazby s vzájemně nespřátelými přáteli. Shlukování je klíčové ke studiu sociálních sítí, protože naznačuje, že dynamické chování sítě není náhodné a že komunikace jedince s novými jedinci je silně ovlivněna informacemi, které mu poskytli již známí jedinci (přátelé).

2.2.8 Modularita (Modularity)

Modularita je široce používaná metrika, která stanovuje, jak dobře lze síť rozdělit na komunity neboli je měřítkem struktury sítě nebo grafu. Je definována jako rozdíl mezi počtem hran, které spadají do komunit a počtem předpokládaných hran v případě náhodného grafu. Formálně je modularita definována vzorcem (6)

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j) \quad (6)$$

kde A je matice sousednosti (A_{ij} je počet hran mezi uzly i a j , není-li mezi nimi žádná vazba $A_{ij} = 0$), k_i a k_j jsou stupně uzlů i a j , m je celkový počet vazeb v síti, c_i a c_j jsou komunity uzlů i a j a v případě, že uzly i a j spadají do stejné komunity ($c_i = c_j$) $\delta(c_i, c_j) = 1$, jinak $\delta(c_i, c_j) = 0$. Příklad matice sousednosti lze vidět v Tab. 1.

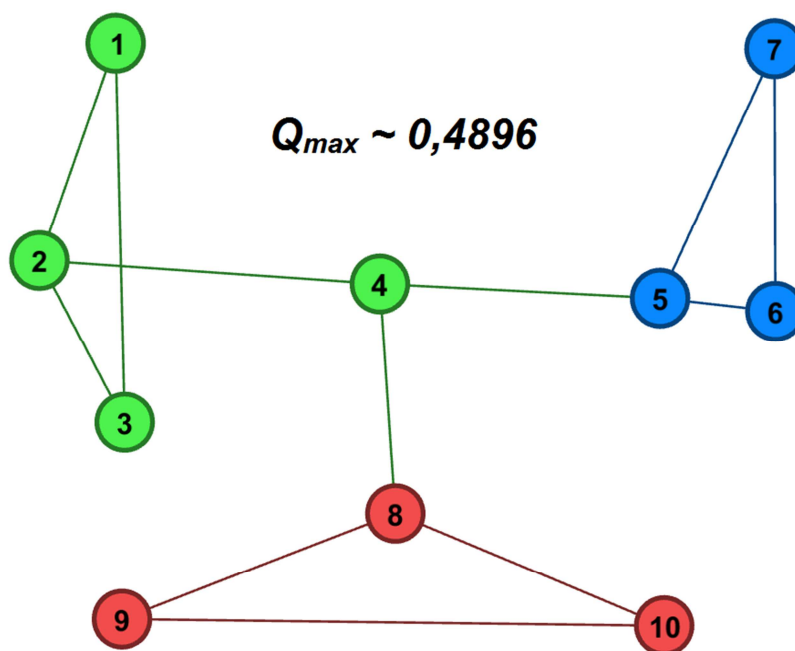
Uzel	1	2	3	4	5	6	7	8	9	10
1	0	1	1	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	1	1	0	0	0	0	0	0	0	0
4	0	1	0	0	1	0	0	1	0	0
5	0	0	0	1	0	1	1	0	0	0
6	0	0	0	0	1	0	1	0	0	0
7	0	0	0	0	1	1	0	0	0	0
8	0	0	0	1	0	0	0	0	1	1
9	0	0	0	0	0	0	0	1	0	1
10	0	0	0	0	0	0	0	1	1	0

Tab. 1: Příklad matice sousednosti pro výpočet modularity

Je velice důležité si uvědomit, že sumace se provádí přes všechny dvojice uzlů i a j v síti tak, jak je znázorněno v Tab. 1, tzn. v úvahu jsou brány jak dvojice uzlů (i, j) , tak dvojice uzlů (j, i) a stejně tak dvojice uzlů v případě, že $i = j$, kdy má uzel vazbu sám na sebe (tzv. self-loop). Podíváme-li se pozorně na vzorec modularity, k nárůstu či poklesu sumy dochází pouze v případě dvojic uzlů, jež spadají do stejné komunity ($c_i = c_j$). V případě dvojice uzlů z různých komunit ($c_i \neq c_j$) je člen sumy vždy funkcí $\delta(c_i, c_j)$ nulován.

Čím vyšší je modularita, tím lépe lze síť rozdělit do komunit. Podle autorů [9] by se hodnota modularity měla pohybovat v intervalu $\langle -1; 1 \rangle$ a dosahuje-li alespoň hodnoty 0,3, vykazuje síť výraznou komunitní strukturu.

Modularita se využívá pro detekci komunit v síti, k čemuž ji využili i autoři v [9]. Abychom mohli vyčíslit hodnotu modularity, musíme znát rozdělení sítě do komunit a smyslem jejich detekce je nalézt takové rozdělení, při kterém dosáhneme nejvyšší hodnoty modularity. Tomu se říká optimalizace nebo maximalizace modularity. Na Obr. 4 lze vidět graf s barevně odlišenými komunitami, který odpovídá matici sousednosti z Tab. 1. Při tomto rozdělení dosahuje síť maximální hodnoty Q .



Obr. 4: Graf s barevně odlišenými komunitami odpovídající matici sousednosti z Tab. 1

2.3 Modely komplexních sítí

Nyní bude pozornost zaměřena na teoretické modely síťové struktury. Podle [12] dělíme tyto modely do tří základních skupin. Jako jeden z nejstarších síťových modelů si uvedeme náhodný graf pánů Solomonoffa a Rapoport (1951) a Erdöse a Rényia (1960). Poté se podíváme na tzv. „model malého světa“ („small-world model“), model struktury sociálních sítí, který byl představen pány Wattsem a Strogatzem (1998), a jako poslední si uvedeme model růstu sítě, který na rozdíl od předchozích dvou, zahrnuje vývoj síťové struktury v čase. Konkrétně se podíváme na model známý jako „bezškálovitý model“ („scale-free model“), který představili pánové Barabási a Albert (1999). Tento model, ve smyslu dynamiky růstu sítě, věrohodně vysvětluje mocninné distribuce stupňů vyskytujících se na World Wide Webu a mnoha dalších sítích.

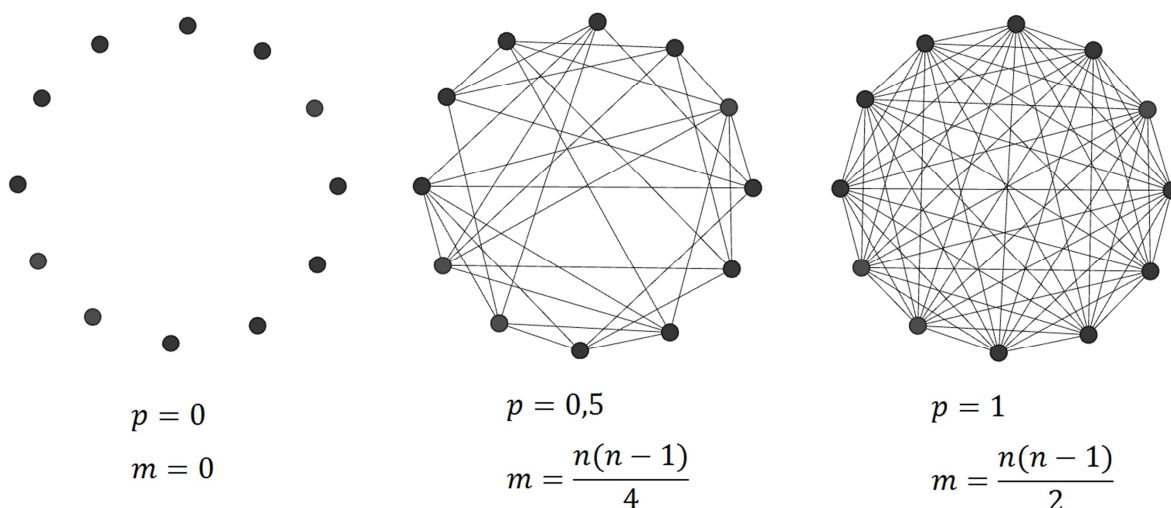
2.3.1 Model náhodného grafu (Random graph model)

Jak je uvedeno v [12], jedním z nejjednodušších a nejstarších síťových modelů je náhodný graf, který představili pánové Solomonoff a Rapoport (1951) a který byl poté intenzivně studován Paulem Erdösem a Alfrédem Rényiem na přelomu padesátých a šedesátých let.

Erdős a Rényi se ve své práci zabývali dvěma rozdílnými modely, přičemž oba označovali jako náhodné. Tyto dva modely jsou běžně odlišeny pomocí symbolických jmen $G_{n,m}$ a $G_{n,p}$:

- $G_{n,m}$ je množina všech možných grafů obsahující n uzlů a m hran. Náhodný graf z této množiny získáme jednoduše tak, že náhodně vybereme m hran ze všech potenciálních $\binom{n}{2}$ neboli $\frac{n(n-1)}{2}$ hran.
- $G_{n,p}$ je množina všech možných grafů obsahujících n uzlů, kde je každá dvojice uzlů propojena vazbou s pravděpodobností p . Náhodný graf z této množiny získáme tak, že procházíme přes všechny $\binom{n}{2}$ neboli $\frac{n(n-1)}{2}$ dvojice uzlů a vkládáme mezi ně vazbu s pravděpodobností p , nebo nevkládáme s pravděpodobností $1-p$.

Oba modely jsou velice podobné s tím rozdílem, že u prvního případu známe přesný počet hran m , u druhého pouze přibližný, který je úměrný pravděpodobnosti p . V případě, že by se pravděpodobnost p proporcionálně rovnala počtu hran m (samozřejmě v případě stejného počtu uzlů n), tedy např. pro $p = 0,5$ a $m = \frac{n(n-1)}{4}$, počet hran v obou grafech by měl být zhruba stejný. Názornou ukázkou je možné vidět na Obr. 5.



Obr. 5: Tři náhodné grafy pro různá p a různá m

Zajímavé je, že téměř všechny náhodné grafy mají stejné vlastnosti, resp. podle [11] se pravděpodobnost stejných vlastností v limitě pro $n \rightarrow \infty$ blíží k 1. Průměrný koeficient shlukování je téměř totožný s pravděpodobností p , což je výrazně méně než u reálných komplexních sítí, a distribuce stupňů je Poissonova, což znamená, jak jsme již uvedli v popisu *Degree distribution*, že existuje určité maximum, kolem kterého se pohybuje

většina uzlů, resp. jejich stupňů. Distribuce stupňů v náhodném grafu tedy nerespektuje mocninný zákon a nemůžeme proto pomocí náhodných grafů nikdy správně modelovat reálné komplexní sítě.

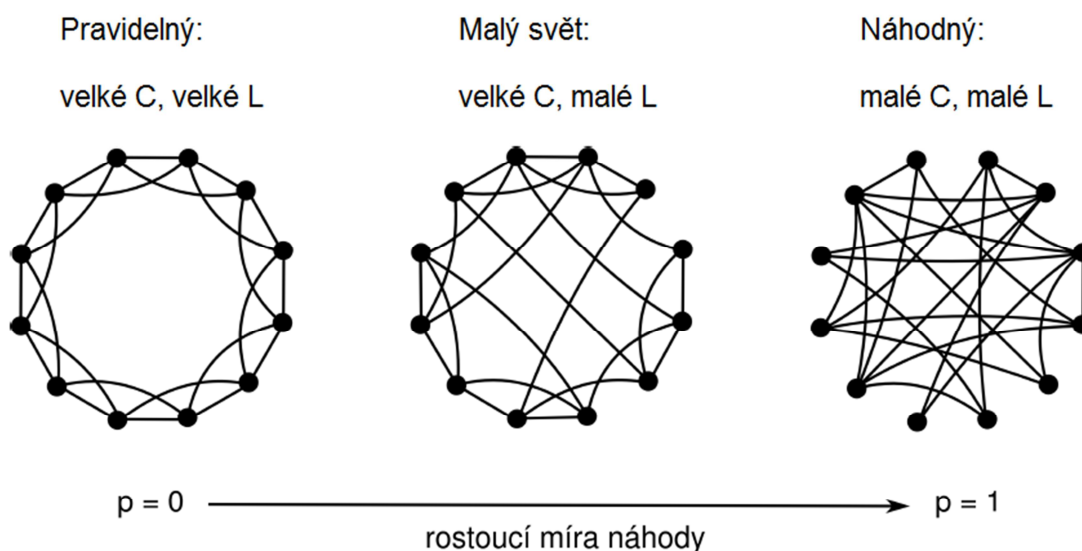
2.3.2 Model malého světa (Small-world model)

Nyní bude pozornost věnována modelu tzv. malého světa, který, jak uvádí [12], představili pánové Watts a Strogatz (1998). Model malého světa (někdy též nazývaný Watts-Strogatz model) je motivovaný faktem, že mnoho reálných světových sítí vykazuje dvě vlastnosti:

- Efekt malého světa (Small-world efekt), což znamená, že většina párů uzlů je v síti dosažitelná na velmi krátkou vzdálenost (více podrobností viz výše).
- Vysoké shlukování nebo tranzitivita (též již podrobně popsáno výše), což znamená, že existuje zvýšená pravděpodobnost toho, že dva uzly budou přímo spojeny hranou v případě, že mají společný sousedící uzel. V řeči sociálních sítí to znamená, že dva lidé se mnohem pravděpodobněji znají v případě, že mají jiného společného známého.

Ale jak vytvořit síť, která by splňovala obě tyto vlastnosti? Odpověď Wattse a Strogatze spočívá v kombinaci pravidelného a náhodného grafu tak, jak lze vidět na Obr. 6. Začneme s pravidelnou mřížkou, kterou vytvoříme tak, že uspořádáme n uzlů do kruhu a každý uzel spojíme s jeho k sousedy ($k/2$ na každé straně). V Obr. 6 platí $k = 4$. V této mřížce procházíme postupně všechny vazby a každou s pravděpodobností p přepojíme do jiného náhodného uzlu. Výsledek je síť, kterou je možné vidět uprostřed Obr. 6, to je small-world model. Ten interpoluje mezi pravidelným a náhodným grafem. Pro $p = 0$ nebude žádná vazba přepojena a výsledek je pravidelný graf s vysokým koeficientem shlukování (C), ale velkými vzdálenostmi mezi uzly (L). Naopak pro $p = 1$ jsou přepojeny všechny vazby a výsledek je téměř ekvivalentní s náhodným grafem, který má krátké vzdálenosti mezi uzly, ale zase nízký koeficient shlukování. Mezi tím však existuje značný rozsah hodnot pro p , při kterém síť splňuje jak vysoké shlukování, tak krátké vzdálenosti.

Model malého světa tedy na rozdíl od modelu náhodného grafu kromě krátkých vzdáleností splňuje i vysoké shlukování. Ovšem stále ještě nerespektuje mocninný zákon v distribuci stupňů, která má stejně jako u náhodného grafu zhruba Poissonovo rozdělení.



Obr. 6: Model malého světa vzniká kombinací pravidelného a náhodného grafu

Zdroj: [11]

2.3.3 Modely scale-free sítí (Scale-free networks)

Jak uvádí autoři v [12], pozorováním distribucí stupňů u reálných sítí bylo zjištěno, že u mnoha systémů, včetně citačních sítí, World Wide Webu, Internetu a metabolických sítí, lze distribuci stupňů aproximovat mocninným zákonem (viz výše). Pro tyto sítě zavedli pánové Barabási a Albert v roce 1999 termín *scale-free networks* (bezškálové sítě). Fakt, že zjevně tak odlišné systémy, jako je World Wide Web, citační a metabolické sítě, sdílí bezškálové rozdělení stupňů je vedlo k myšlence, že struktura všech těchto sítí je tvořena společným mechanismem.

Mechanismus, který Barabási a Albert navrhli, měl dvě složky. Za prvé, síť roste, uzly jsou k ní postupně přidávány, neboli dynamicky se mění v čase. Tento fakt nerespektoval model náhodného grafu ani model malého světa, kde byla síť považována za statickou s pevným počtem uzlů. A za druhé uzly získávají nové vazby proporcionálně počtu vazeb, které již mají. To je proces, kdy uzly v rámci vytváření nových vazeb preferují uzly s vyšším stupněm (degree), to znamená, je pravděpodobnější, že jedinec naváže kontakt s někým, kdo disponuje mnoha vazbami (např. má mnoho přátel), než s někým, kdo jich má málo. Lze se setkat i s frází „*rich get richer*“ („*bohatí bohatnou*“), která je základem tohoto procesu a přesně vystihuje jeho podstatu. Tento proces byl Barabásim a Albertem nazván jako *preferential attachment* neboli upřednostňované připojování.

První složka je nepochybně pravdivá. Všechny studované sítě v průběhu let neustále rostly. I druhá složka se jeví jako velmi logická. Např. webové stránky se stávají velmi

známými díky odkazům z jiných stránek a je velmi pravděpodobné, že na velmi známé stránky budou vznikat další odkazy.

Růst a upřednostňované připojování poskytují jednoduché vysvětlení vzniku bezškálových sítí. Pro demonstraci navrhli Barabási a Albert model (označovaný jako Barabási-Albert model), ve kterém síť v každém časovém kroku roste přidáním jednoho uzlu, který je připojen m vazbami, přičemž uzly, ke kterým je připojen, jsou vybírány náhodně s pravděpodobností úměrnou jejich stupni. Formálně *preferential attachment* definujeme následovně. Necht' k_i je stupeň uzlu i . Potom pravděpodobnost, že se nově přichodzí uzel propojí s uzlem i definujeme vzorcem (7)

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (7)$$

neboli jako poměr stupně uzlu i a součtu stupňů všech uzlů.

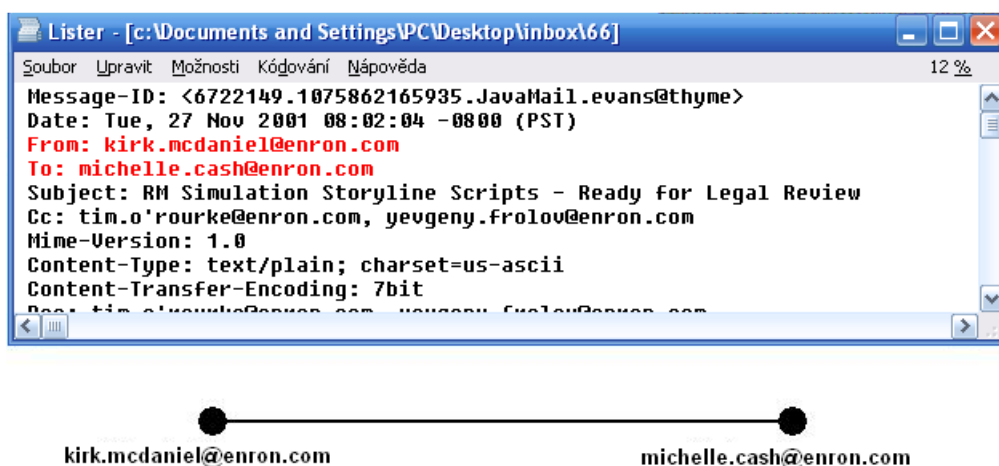
Jak je uvedeno v [12], bylo vytvořeno mnoho verzí Barabási-Albert modelu a mnoho variant upřednostňovaného připojování, ovšem jako základní model *scale-free* sítě je považován ten, který jsme si popsali výše.

3 Dynamika sociálních sítí

Většina studií zabývajících se analýzou sociálních sítí nebo obecně sociálními sítěmi považovala tyto sítě za statické neboli neměnné v čase. Nemusíme se příliš zamýšlet nad tím, že to není úplně pravda a že v rámci sociálních sítí právě čas hraje velmi významnou, neřkuli nejvýznamnější roli. Sociální sítě jsou jednoduše velmi proměnlivé a vyvíjí se v čase. Tím se dostáváme k problematice, která je hlavním zájmem celé této diplomové práce – dynamika sociálních sítí.

3.1 Pojem dynamiky a její význam

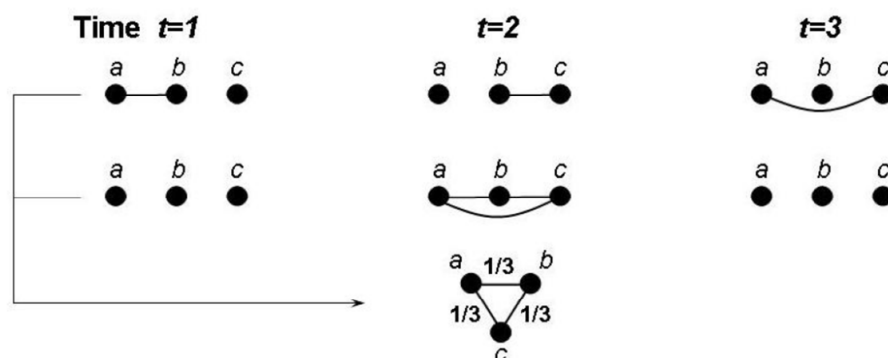
Tradiční techniky analýzy sociálních sítí se na dynamiku příliš nezaměřují. Dynamika je zanedbána i v provedené analýze ve výše zmíněné bakalářské práci. Vezměme si v [3] prezentovanou síť sestavenou na základě e-mailové korespondence, konkrétně na základě půl milionu e-mailů. Každý poslaný e-mail představuje jednu vazbu mezi dvěma uzly. Příklad zpracování jednoho e-mailu lze vidět na Obr. 7.



Obr. 7: Příklad e-mailu a jeho zpracování

V úvahu bylo bráno pouze kdo, komu (řádek *From:* a *To:*) a kolik e-mailů poslal v průběhu určitého časového období. To, kdy byly jednotlivé e-maily odeslány, bylo zcela zanedbáno, i když tuhle informaci máme zjevně také k dispozici (řádek *Date:*). Tímto způsobem byla síť zachycena v jednom momentálním okamžiku a to na konci sledovaného časového období. V případě, že by jednotlivé e-maily byly doplněny o informaci o tom, kdy došlo k jejich odeslání, získali bychom možnost sledovat vývoj sítě po celou dobu její existence, z čehož by zřejmě bylo možné vytěžit nemálo informací o této síti.

Uveďme si jeden velice jednoduchý příklad demonstrující důležitost dynamiky neboli to, co díky ní můžeme zjistit. Obr. 8 ilustruje, že v některých případech mohou velmi odlišná dynamická data vést ke stejnému statickému grafu, z čehož vyplývá, že rozhodnutí založená pouze na statických datech mohou být chybná.



Obr. 8: Dva dynamické grafy, které se mapují na stejný statický graf

Zdroj: [13]

Předpokládejme, že hrany v Obr. 8 představují sociální kontakt, který může způsobit přenos nemoci. My máme možnost naočkovat pouze jednoho člověka s cílem minimalizovat počet nakažených na konci s tím, že k přenosu nemoci dojde při každém kontaktu a každý jedinec může být nakažen již na počátku. V případě statického grafu se zdá, že bez ohledu na to, který jedinec je naočkován, budou na konci nakaženi dva jedinci. Nicméně v prvním dynamickém grafu za předpokladu, že k počáteční nákaze dochází v kroku 0, je potřeba naočkovat b . V tom případě se nemoc nerozšíří bez ohledu na to, zda je počátečně nakaženo a , nebo c . Respektive k přenosu nemoci může dojít právě mezi a a c , ale až v kroku 3. Reprezentace statickým grafem nám tudíž znemožňuje ptát se na to, jak rychle se může nemoc šířit populací a kteří jedinci by měli být naočkováni, aby došlo ke zpomalení šíření nemoci.

3.2 Aktuální studie v dynamice sociálních sítí

Dynamika sociálních sítí je ovlivňována množstvím vzájemně provázaných procesů. Hlubší pochopení těchto procesů nám umožňuje modelovat a předpovídat strukturu a dynamiku v sociálních sítích. To má velké praktické využití při návrhu infrastruktury, aplikací a bezpečnostních mechanismů sociálních sítí.

I přes tradiční statickou analýzu existuje mnoho studií, které se zabývají analýzou a modelováním dynamiky a snaží se předpovídat chování jedinců v sociálních sítích. Pochopení vlastností vztahů mezi uživateli sociálních sítí je totiž předpokladem k interpretaci sociální dynamiky. [4]

Podle [13] výzkum v analýze dynamiky sítí postupoval v několika směrech. Podle statistické mechaniky platí, že sítě jsou komplexní fyzické systémy a tato mechanika se snaží popsat zákony jejich vývoje. Více výpočetní pohled zahrnuje pravděpodobnosti a nejistotu do strukturovaných informací a kombinuje modelování sociálních sítí s multiagentními systémy. Počítačové simulace byly donedávna hlavní výpočetní technikou zahrnující dynamické informace o síti.

S příchodem online sociálních sítí máme k dispozici rozsáhlá data a topologie sítě byla doplněna o homofilní rysy (popsány výše), které modelují podobnost uživatelů v síti. Nicméně, sociální blízkost mezi uživateli v reálném světě není závislá pouze na online sociálních sítích, ale také na pohybu uživatelů. Dostupnost těchto dat nám s příchodem GPS telefonů a *location-based* sociálních sítí (sociálních sítí založených na poloze) poskytuje nový zdroj informací o uživateli.

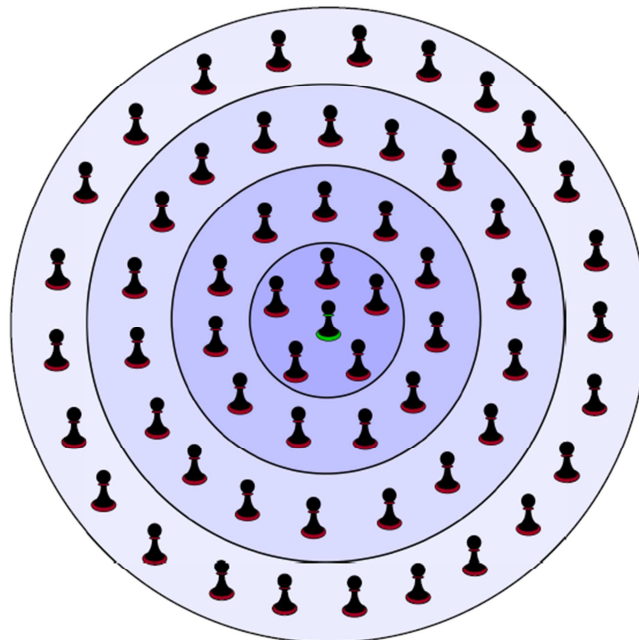
Autoři v [4] se snažili předpovídat partnerství dvou uživatelů právě za pomoci kombinace dat z online sociální sítě a dat z *location-based* sociální sítě, která byla získána z virtuálního světa Second Life. Byly zjištěny významné rozdíly mezi dvojicemi uživatelů, kteří jsou ve vztahu a dvojicemi, které ve vztahu nejsou. K modelování sociální blízkosti byla analyzována topologie sítě (např. počet společných kontaktů atd.) společně s homofilními rysy (návštěva společných akcí, vzdálenost mezi uživateli) získanými z *location-based* sociální sítě. Právě tyto homofilní rysy byly identifikovány jako nejcennější z pohledu předpovědi partnerství. Analýza obrovského množství rysů sociální blízkosti jak z online sociální sítě, tak z *location-based* sociální sítě umožňuje předpovídat partnerství mezi uživateli s přesností 93,3 %, kdy experimenty ukázaly, že kombinace rysů z obou těchto domén (online SS a *location-based* SS) překonává rysy obou z nich.

Vzhledem k dramatickému růstu dostupných dat z online sociálních sítí společně s možností automatizace sběru těchto dat vzrůstá zájem také o analýzu síťové dynamiky, která se zaměřuje na vývoj technik speciálně navrhovaných pro velmi rozsáhlé a dynamické sítě. Například výpočet centrality založený na nejkratších cestách vyžaduje náročné nalezení nejkratších cest mezi všemi páry uzlů. Vzhledem k tomu, že běžné způsoby výpočtu centrality byly navrženy pro statické sítě, pokus o takový výpočet v rámci dynamické sítě by se zúžil pouze na její momentální stav a při dynamické změně sítě by se musel opakovat.

V [7] proto zaměřují pozornost na postupný (inkrementální) výpočet nejkratších cest v rámci analýzy síťové dynamiky a diskutují možné vylepšení tradičních technik analýzy sociálních sítí pomocí technik inkrementálních algoritmů, což jsou speciálně navržené algoritmy pro dynamicky se měnící sítě, které reagují na změny analyzované sítě v průběhu času. Při jakékoliv změně sítě nejprve identifikují části, které byly změnou ovlivněny a aktualizují pouze je. Jednoduše tyto algoritmy šetří redundantní výpočty, jež byly provedeny již dříve.

Autoři v [7] se takto snaží zjednodušit opakovaný výpočet *closeness centrality* (více v [3]) při dynamické změně sítě. Výkon navrhovaného algoritmu se logicky zvyšuje s velikostí sítě. Očekávaným poznatkem může být to, že je-li průměrná délka nejkratších cest v síti nízká, celková hloubka stromu nejkratších cest je menší, což ve výsledku znamená méně aktualizací při potřebě přepočítat nejkratší cesty v síti.

Odlišný pohled na studium sociálních sítí byl přijat antropology a evolučními psychology. Spíše než na globální vlastnosti sítě se zaměřili na lokální vlastnosti osobních sociálních sítí, často nazývaných ego sítě (ego networks). Ego síť je jednoduchý model popisující sociální vztahy mezi individuem (ego) a všemi kontakty, které toto individuum má (angl. alters). Model ego sítě je možné vidět na Obr. 9.



Obr. 9: Model ego sítě

Zdroj: [1]

Nejdůležitějším poznatkem v rámci ego sítí je, že existuje určitý limit kontaktů (alters), které je jedinec schopen aktivně udržovat vzhledem ke kognitivnímu (poznávacímu) omezení lidského mozku. Jak je uvedeno v [1], tento limit objevil Robin Dunbar, který studoval primáty a našel vzájemný vztah mezi velikostí mozkové kůry (část mozku zodpovědná za sociální aktivitu) u různých lidoopů a velikostí jejich sociálních skupin. Dunbar předpověděl, že velikost sociálních skupin u lidí, s ohledem na velikost jejich mozku, by měla být okolo 150 – toto číslo je známo jako Dunbarovo číslo. Dunbarova hypotéza byla validována mnoha experimenty jak na offline, tak na online sociálních sítích.

Další důležitý poznatek o ego sítích je, že kontakty kolem ega vytváří strukturu vrstev v podobě soustředných kruhů, které mají typické vlastnosti a velikost. Podle [1] jsou v antropologii a psychologii v ego síti běžně identifikovány čtyři vrstvy. První z nich, nazývaná *support clique*, zahrnuje kontakty, u kterých jedinec hledá pomoc v případě silné emoční tísně nebo finanční krize a průměrně obsahuje 5 členů. Další vrstvy se nazývají *sympathy group* (15 členů), *affinity group* (50 členů) a *active network* (150 členů). Poslední vrstva koresponduje s počtem lidí, se kterými je jedinec schopen aktivně udržovat kontakt, který určuje Dunbarovo číslo. Pozoruhodná vlastnost těchto vrstev je, že poměr velikosti přilehlých vrstev se zdá být konstantní a je blízký hodnotě 3.

Většina předchozích studií se zaměřovala na zachycení dynamiky do jednoho jednoduchého procesu. Snahou autorů v [9] je zachytit dynamiku z různých pohledů, a to z pohledu jednotlivce, komunit a celé sítě, kdy byla analyzována čínská sociální síť, a bylo zjištěno, že uživatelé vytváří nejvíce vazeb krátce po připojení do sítě, ovšem s postupným růstem sítě dominují ve vytváření nových vazeb již existující uzly. Vliv *preferential attachmentu* v průběhu času slábne, což může reflektovat sníženou viditelnost uzlů s růstem sítě. K pochopení toho, jaký dopad na aktivitu uživatele mají komunity, bylo autory v [9] porovnáno vytváření nových vazeb mezi uživateli v komunitách a uživateli mimo jakoukoliv komunitu. Výsledky ukázaly, že uživatelé, kteří se nacházejí v rozsáhlých komunitách, jsou více aktivní v hledání přátel a zůstávají aktivní po delší dobu. Vykazují tedy vyšší aktivitu ve všech rozměrech, což potvrzuje pozitivní vliv komunit na uživatele.

Komunity v sociálních sítích mohou být na základě struktury sítě definovány jako skupiny hustě propojených uzlů. Přesněji řečeno jsou to skupiny, kde více vazeb spojuje uzly v rámci jedné komunity než vazeb mezi odlišnými komunitami. Komunity se implicitně odvíjejí od struktury sítě, tzn., uzly se nemohou připojit anebo naopak komunitu opustit. Komunity efektivně zachycují „sousedství“ v sociální síti.

V [9] věří v to, že komunity představují nejlepší abstrakci pomocí níž lze měřit vliv sociálního sousedství na dynamiku jednotlivce. V článku byly použity čtyři metriky k identifikaci strukturálních změn v obrovské čínské sociální síti a to *average degree*, *average path length*, *average clustering coefficient* a *assortativity* (popsány výše). Všechny tyto ukazatele se stabilizovaly během prvních 100 dní.

Nedávný výzkum se podle [10] zabýval i komplexnějšími modely sociálních sítí než těmi studovanými v minulosti, které se většinou zabývaly základními mechanismy růstu sociální sítě. Novější výzkumy se oproti tomu zaměřují na pokročilé funkce jako je důvěra, doporučení, kooperace a výměna informací. V [14] zase přišli s myšlenkou kompozitních (složených) sítí. Kompozitní (složenou) síť je myšleno spojení dvou nebo více individuálních sítí do jedné (např. Facebook a LinkedIn). Uživatelé, kteří mají účty na obou těchto sítích, mohou fungovat jako mosty (bridges) mezi těmito sítěmi, tzn., že jedna individuální síť může ovlivňovat dění v té druhé. Modelování dynamiky pouze na základě jedné individuální sítě může být značně nekompletní a autoři věří, že modelování dynamiky kompozitních sítí může vést k přesnějším a komplexnějším výsledkům a může pomoci pochopit interakce mezi rozdílnými sítěmi. Nicméně, nelze považovat dvě individuální spojené sítě za jednu (např. Facebook + Twitter), protože každá z těchto sítí má jiné vlastnosti. Facebook reflektuje uživatelské každodenní vazby, kdežto Twitter je zaměřen spíše na získávání aktuálních novinek.

3.2.1 Dostupnost reálných dat

Co se týká dostupnosti dat z reálných sociálních sítí, mnoho datasetů můžeme nalézt v [15], ovšem nejlépe by se tyto procesy zkoumaly v rámci dnešních obrovských online sociálních sítí jako je Facebook, LinkedIn atd. Bohužel poskytovatelé těchto sociálních sítí obecně považují dynamická data za obchodní tajemství a mají málo podnětů k tomu, aby tato data zpřístupnili výzkumu. Místo toho studie analyzují citační grafy, sítě se sdíleným obsahem a vysokou úroveň statistiky sociálních sítí. Jiné studie se snaží validovat generativní modely jako je například *preferential attachment* [9] nebo je naopak snahou vytvářet nové simulační modely, které budou poskytovat výstupy srovnatelné s těmi z reálných sociálních sítí.

I přes obecnou nedostupnost dat byla provedena zajímavá studie na Twitteru. V [1] analyzovali dataset obsahující komunikaci více než 2 milionů uživatelů. K analýze byl použit právě model ego sítí, který byl popsán výše. Výsledky odhalily to, že lidé na Twitteru se chovají úplně jinak než na ostatních sociálních sítích. Klíčovým zjištěním je, že uživatelé

mají málo silných vazeb, pokud vůbec nějaké. Místo toho disponují množstvím slabých vazeb, které se velice rychle obměňují. Toto chování dobře zapadá do naší extrémně dynamické společnosti, kde se lidé musí rychle adaptovat na frekventované změny v jejich životě od citové sféry až po jejich zaměstnání.

3.3 Metodiky modelování a zkoumání dynamiky

V této části práce si uvedeme dva příklady metodik modelování, a to systémovou dynamiku a agentové modelování. Poté se podíváme, jakým způsobem můžeme zkoumat dynamiku sociálních sítí. Řekneme si něco o snímkování sítě, od něhož se dostaneme k teorii časových řad a jejímu praktickému uplatnění.

3.3.1 Systémová dynamika (System Dynamics)

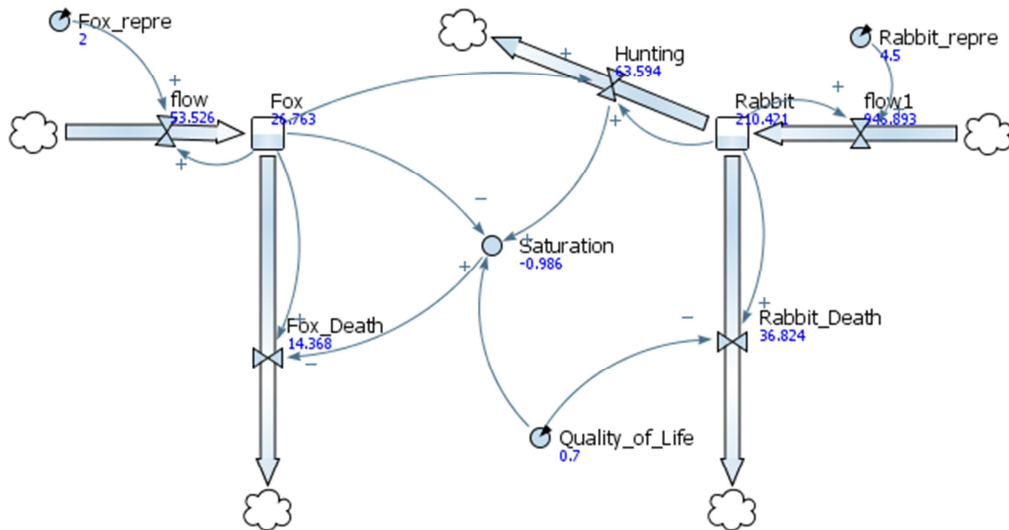
Systémová dynamika je perspektivní sada nástrojů, které nám umožňují pochopit strukturu a dynamiku komplexních systémů. Systémová dynamika je také způsob modelování, který nám umožňuje vytvářet formální počítačové simulace komplexních systémů.

Metodika systémové dynamiky (SD) se obvykle používá u dlouhodobých, strategických modelů a předpokládá vysokou úroveň agregace objektů, jež jsou modelovány. Lidé, výrobky, události a další diskrétní položky jsou v modelech zastoupeny podle jejich množství, takže ztrácí všechny individuální vlastnosti, historii nebo dynamiku. Pokud je tato úroveň abstrakce vhodná pro řešený problém, SD může být tou správnou metodou.

Příkladem by mohlo být modelování dynamiky počtu klientů jednotlivých mobilních operátorů. Systémová dynamika pracuje s tzv. zásobárnami a toky mezi těmito zásobárnami. V tomto případě by zásobárny tvořili klienti jednotlivých mobilních operátorů, kde každý operátor může přilákat nové klienty anebo přetáhnout část klientů od konkurence. Pohyby klientů tvoří zmíněné toky. Tyto toky jsou samozřejmě ovlivňovány mnoha faktory, které je možno modelovat.

Jiný příklad může být dynamika populace lišek a zajíců v přírodě. Jak lišky, tak zajíci se určitou rychlostí reprodukují a umírají. Kromě toho zajíci tvoří potravu pro lišky, tzn. s rostoucím počtem lišek, klesá počet zajíců. Ovšem je-li málo zajíců, není dostatek potravy pro všechny lišky a ty tím pádem umírají. Náhled tohoto modelu v simulačním prostředí AnyLogic lze vidět na Obr. 10. V tomto případě nás individuální charakteristiky lišek a zajíců nezajímají, zajímá nás pouze jejich počet.

Pokud však jsou pro nás individuální údaje důležité, je vhodnější využít následující metodiku, a tou je agentové modelování.



Obr. 10: Modelování populace lišek a zajíců metodou systémové dynamiky

3.3.2 Agentové modelování (Agent Based Modeling)

I když je v literatuře možné najít celou řadu různých definic agentového modelování (ABM), z hlediska praktických aplikací lze agentové modelování definovat jako v podstatě decentralizovaný, individuálně orientovaný (oproti úrovni systému) přístup k návrhu modelu. Při navrhování takového modelu jsou nejprve identifikovány aktivní subjekty, tzv. agenti, což mohou být lidé, firmy, projekty, vozidla, města, zvířata, lodě, výrobky, případně další, a je definováno jejich chování prostřednictvím vlivů, reakcí, paměti, stavů atd. Takto navržené agenti jsou poté vloženi do určitého prostředí, je stanoveno propojení mezi nimi a je spuštěna simulace. Globální chování se pak projevuje v důsledku interakcí mnoha jednotlivých chování.

Tato metodika je ze své podstaty ideální pro modelování sociálních sítí, kdy modelujeme individuální chování jednoho člověka a poté sledujeme globální chování celé sítě. Individuální jedinci v síti vystupují v modelu právě jako výše zmiňovaní agenti. Model, kterým se budeme zabývat v praktické části, je, intuitivně, tvořen právě touto metodikou.

Více informací a detailní popis zmíněných metodik modelování můžeme nalézt např. v [16].

3.3.3 Snímkování sítě

Nyní se pojděme podívat na to, jakými metodami lze zkoumat dynamiku sítí. Autoři v [1] k tomuto zkoumání použili snímky sítě, tzv. snapshoty, což není nic jiného než zachycení stavu sítě v určitém okamžiku do statického grafu, který je následně možné zkoumat s využitím tradičních metod analýzy sociálních sítí (viz výše). Zkoumáním jednoho statického grafu můžeme zjistit jistě hodně, ovšem zjištěné údaje se vztahují pouze k jednomu časovému okamžiku, chybí nám tedy ona dynamika. O té z jednoho snímku nevíme vůbec nic. Dynamiku sítě začneme odhalovat až v momentě, kdy jednotlivé snímky budeme mezi sebou porovnávat.

Uvážíme-li, že bude naším cílem zkoumat dynamiku nějaké sítě v průběhu jednoho roku, máme možnost vytvořit jeden snímek například každý den, čímž získáme 365 statických snímků sítě seřazených od minulosti do přítomnosti. Porovnáme-li dva po sobě jdoucí snímky, získáme představu o změnách v síti během jednoho dne, porovnáme-li např. první a třicátý, získáme představu o změnách během jednoho měsíce a porovnáme-li první snímek s posledním, získáme představu o vývoji sítě v rámci celého roku. Velikost intervalu mezi jednotlivými snímky lze samozřejmě dle potřeb upravit. Čím menší bude interval, tím detailnější budou výsledky.

Jak jsme již řekli, snímky lze analyzovat jako jednotlivé grafy, kdy máme možnost zkoumat a měřit jejich vlastnosti. Budeme-li měřit určitou vlastnost grafu v každém snímku sítě, získáme ve výše popisovaném případě 365 hodnot charakterizujících určitou vlastnost pro každý jednotlivý snímek. Tím se dostáváme k oblasti, která je vhodná pro zkoumání takovýchto hodnot, a tou je analýza časových řad.

3.3.4 Analýza časových řad

Jak bylo naznačeno výše, budeme-li zaznamenávat statické ukazatele sítě v průběhu času, nabízí se pro zkoumání jejich vývoje problematika časových řad a jejich analýza.

Časové řady (dynamické řady, vývojové řady) jsou posloupnosti pozorování kvantitativní charakteristiky uspořádaná v čase od minulosti do přítomnosti, pomocí nichž můžeme zkoumat dynamiku jevů v čase. Mají základní význam jak pro analýzu příčin, které na tyto jevy působily a ovlivňovaly jejich chování v minulosti, tak pro předvídání jejich budoucího vývoje.

Při klasické analýze časových řad se vychází z předpokladu, že každá časová řada může obsahovat čtyři složky:

- **trend** - je obecná tendence vývoje zkoumaného jevu za dlouhé období, je výsledkem dlouhodobých a stálých procesů
- **sezónní složka** - je pravidelně se opakující odchylka od trendové složky, perioda této složky je menší než celková velikost sledovaného období
- **cyklická složka** - udává kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje (používáno spíše v makroekonomických úvahách)
- **náhodná (stochastická) složka** - se nedá popsat žádnou funkcí času, "zbývá" po vyloučení trendu, sezónní a cyklické složky

Nejčastěji se při analýze časové řady předpokládá aditivní model popisu chování řady. Předpokládá se, že jednotlivé složky vývoje se sčítají podle (8)

$$y_t = T_t + S_t + C_t + \varepsilon_t \quad (8)$$

kde na pravé straně po řadě vystupují složky trendová, sezónní, cyklická a náhodná. Analýza trendové složky je zřejmě nejdůležitější částí analýzy časových řad. Určení trendu neboli očištění (vyhlazení) časové řady od periodického a nahodilého kolísání můžeme provést dvěma způsoby:

- **mechanicky** – klouzavý průměr
- **analyticky** – trendová funkce

Vyhlazování časové řady klouzavým průměrem spočívá v nahrazování hodnot časové řady průměrem z p po sobě jdoucích hodnot. Často se pro p volí liché číslo, tedy 3, 5, 7, atd. Čím větší je p , tím dochází k většímu vyhlazení, ale roste počet nevyrovnaných hodnot, pro které v důsledku velkého p průměr nelze vypočítat.

Analytické vyhlazení časové řady spočívá v nalezení matematické funkce, která bude trend časové řady popisovat. V průběhu let se potvrdilo, že při výběru trendových funkcí většinou vystačíme s úzkou nabídkou. Nejčastěji používané jsou lineární, kvadratická, exponenciální, logaritmická nebo hyperbolická.

Míry dynamiky

Pro odhalování dynamiky vývoje časových řad lze využít následující míry:

Absolutní přírůstek, získán podle (9)

$$\Delta y_t = y_t - y_{t-1}, t = 2, 3, \dots, n \quad (9)$$

Průměrný absolutní přírůstek, získán podle (10)

$$\bar{\Delta} = \frac{\sum \Delta y_t}{n-1} = \frac{(y_2 - y_1) + (y_3 - y_2) + \dots + (y_n - y_{n-1})}{n-1} = \frac{y_n - y_1}{n-1} \quad (10)$$

Relativní přírůstek, získán podle (11)

$$\delta_t = \frac{\Delta y_t}{y_{t-1}} = \frac{y_t - y_{t-1}}{y_{t-1}} = \frac{y_t}{y_{t-1}} - 1 \quad (11)$$

Průměrný koeficient růstu, získán podle (12)

$$\bar{k} = \sqrt[n-1]{k_1 k_2 \dots k_n} = \sqrt[n-1]{\frac{y_2}{y_1} \frac{y_3}{y_2} \frac{y_4}{y_3} \dots \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}} \quad (12)$$

Směrnici tečny grafu funkce, respektive její předpis, získán podle (13), kde $f'(x_0)$ je derivace funkce v bodě T .

$$y - y_0 = f'(x_0)(x - x_0), T = [x_0; y_0] \quad (13)$$

Detailnější popis problematiky časových řad lze nalézt např. zde [17].

4 Použitý model

V předchozí (teoretické) části práce jsme získali představu o problematice, které se věnujeme a v následující (praktické) části práce se pokusíme demonstrovat její užití v praxi. V další části naší práce si klademe za cíl rozvinout a zdokonalit již dříve navržený model pro simulaci dynamiky šíření informací a znalostí v sociální síti a na experimentálních výsledcích demonstrovat vliv provedených změn i některé další jevy, které se v dynamice sítí projevují.

Model může být využit pro zkoumání dynamiky v sítích různého rozsahu i zaměření vytvořených či používaných za účelem přenosu a šíření informací a znalostí (např. podnikové sítě, online služby, atd.), jeho využití se tedy neomezuje pouze na online služby a čistě elektronický přenos informací.

4.1 Popis modelu

Další část práce navazuje na model popsany v [18] a dále jej rozšiřuje a zdokonaluje. Základem modelu je AB přístup vycházející z koexistence a vzájemné komunikace elementárních objektů (agentů). Model je zaměřen na detailní zkoumání chování agentů v procesu získávání potřebných znalostí a na parametry, které toto chování ovlivňují.

Hlavním východiskem modelu je existence daného počtu agentů v daném prostoru, kteří jsou vystavováni "životním" situacím, vyžadujícím jejich rozhodnutí (řešení situace). Tyto situace mohou mít různou podobu a lze si je též představit jako zprávy, které agenti dostávají a na které musí co nejlépe reagovat. Agenti se snaží o co nejlepší reakci, kvalita každé reakce je v intervalu $\langle 0;1 \rangle$, hodnota 1 odpovídá nejlepší reakci. Při hledání co nejlepší reakce vycházejí agenti z vlastních informací a schopností, ale snaží se zkoumat i řešení (reakce) téže situace přijatá v minulosti ostatními agenty, se kterými proto komunikují. Model vychází z předpokladu uzavřeného světa uplatněného na počet uzlů v síti i velikost množiny možných situací, které jsou v něm konstantní a v čase neměnné.

Každý agent je vybaven pamětí, která obsahuje přijatá řešení událostí, která lze znovu přijmout v případě vystavení stejné události. K tomu každý agent disponuje ještě seznamem agentů, řekněme partnerů, jejichž řešení uložená ve své paměti může taktéž přejímat. Jelikož se agenti, jak jsme již uvedli, snaží přijímat co nejlepší řešení, tak právě zde dochází k šíření informací a znalostí v síti. Při přejímání řešení od partnerů je ovšem třeba brát v úvahu komunikativnost agentů neboli jejich chuť ke komunikaci a zároveň jejich ochotu odpovídat na dotazy ze sítě. To znamená, že vždy agenti chtějí komunikovat a ne

vždy partneri agenta sdílí svá vlastní řešení. V takovém případě, tzn., nenalezne-li agent řešení dané události ve své paměti a nezíská-li žádné řešení ani od jednoho svého partnera, generuje agent na základě svých vlastních schopností neboli inteligence, která je též modelována, své vlastní řešení. Jednoduše vyřeší situaci po svém, přičemž kvalita generovaného řešení je závislá na inteligenci agenta.

Můžeme ještě podotknout, že přijme-li agent nějaké řešení, tak autor tohoto řešení se stává novým partnerem agenta. Velikost seznamu partnerů může být ovšem omezena. Každý agentův partner v seznamu je ještě ohodnocen podle toho, jak kvalitní řešení dodává a v případě překročení kapacity seznamu jsou odstraňováni ti nejhorší z nich. Stejně tak samotná řešení nevydrží v pamětech agentů věčně. Je implementován i proces zapomínání jedinců, to znamená, nevyužije-li agent řešení ve své paměti po určitou dobu, která je modelována individuálně každému agentu (každý agent zapomíná jinak), je řešení z jeho paměti odstraněno, neboli agent ho zapomene. Zde tím pádem dochází ke ztrátě nebo úpadku znalostí a informací v síti. Detailní popis modelu lze nalézt v [18].

Lze říci, že výše uvedený model dobře popisuje sociální síť, jejímž základním cílem je distribuce a sdílení znalostí (o řešení příslušných situací) a mechanismy v této síti. Z provedených experimentů se však ukázalo, že některé algoritmy chování nejsou ideálně nastaveny či zkreslují chování pozorovatelné na reálných sítích tohoto zaměření. Proto byla provedena revize modelu, jejíž výsledky jsou popsány v následující kapitole.

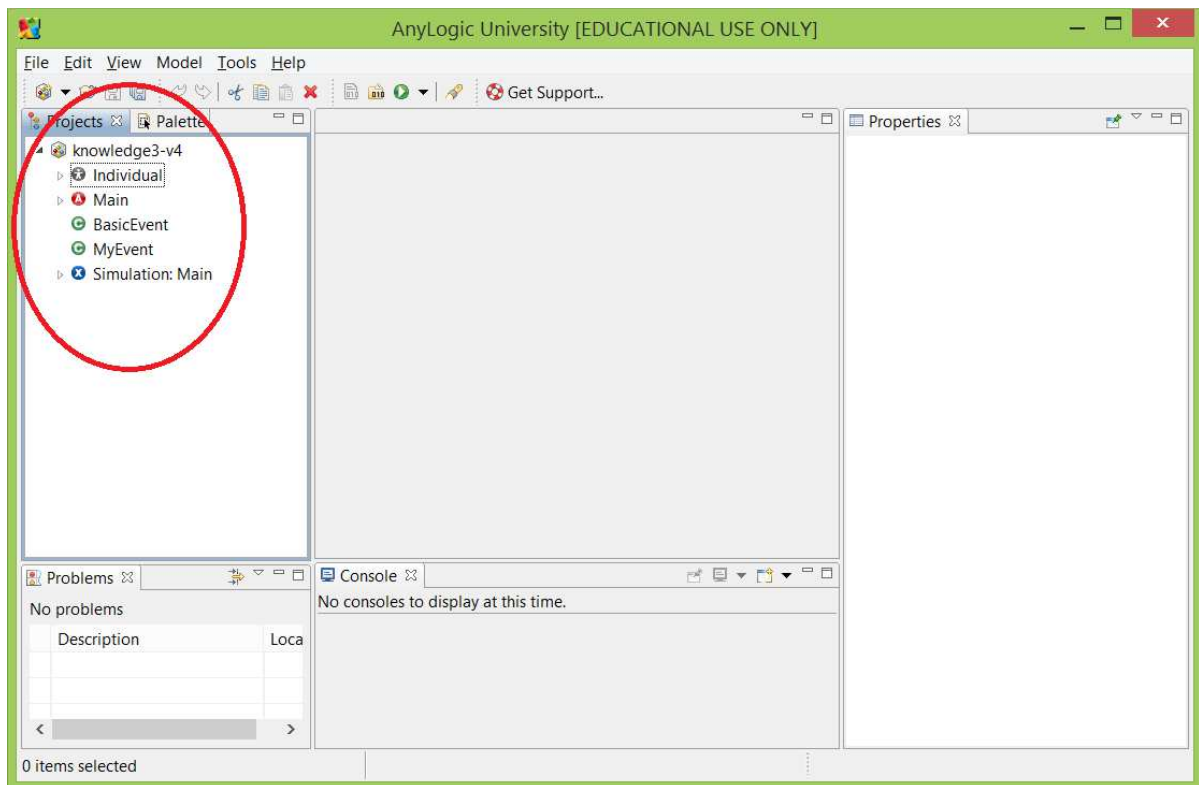
5 Úpravy modelu

Jak již bylo uvedeno, původní model znalostní sociální sítě, který byl k dispozici, je funkční a poskytuje použitelné výstupy pro zkoumání dynamiky sociálních sítí. Tento model byl v rámci této práce dále rozvinut, a to ve dvou základních směrech. Prvním z těchto směrů bylo vylepšení vnitřních mechanismů chování agenta, zejména v oblasti komunikace s ostatními partnery v síti. Cílem bylo posílit využívání vlastní inteligence agentů, kteří spíše preferovali získávání znalostí ze sítě. Druhou oblastí byla úprava celého modelu, který byl konstruován s předpokladem tzv. uzavřeného světa, kdy nebylo možné měnit počet situací, jimž byli agenti vystavováni, ani počet těchto agentů. Do modelu byl proto implementován mechanismus rostoucí sítě s upřednostňovaným připojováním. V následujících částech práce se se všemi úpravami a zásahy do stávajícího modelu postupně seznámíme.

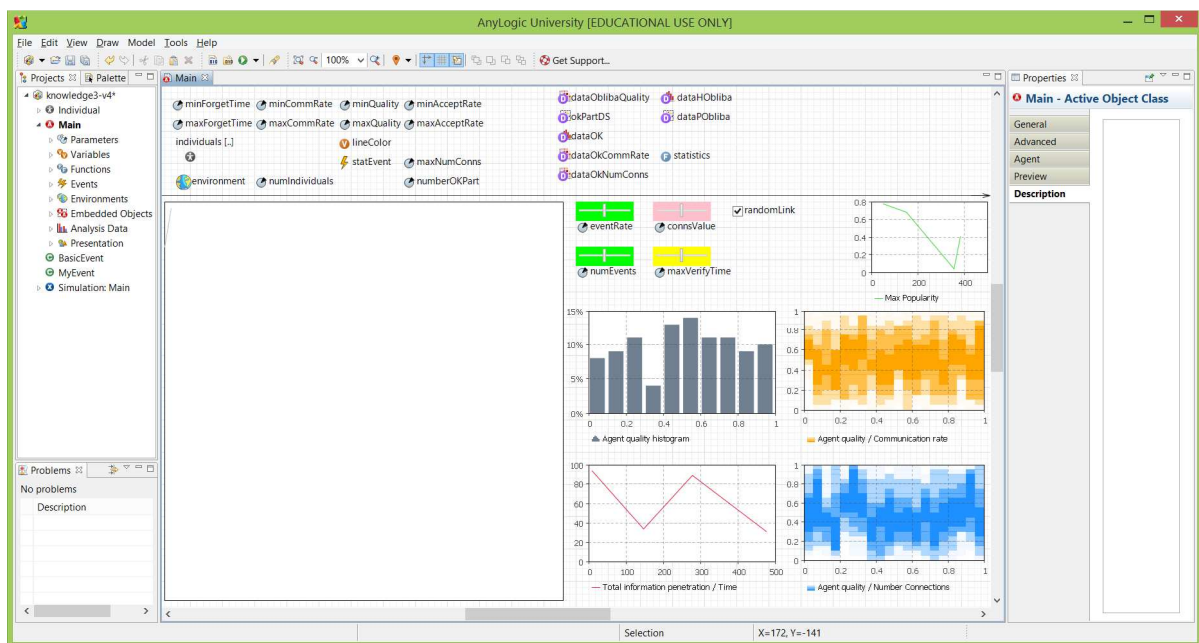
5.1 Použitý software a transformace do nové verze

Model, se kterým budeme pracovat, byl původně implementován v simulačním softwaru AnyLogic, a to konkrétně ve verzi 6.9.0. Jedná se o pracovní verzi modelu pro účely výzkumu, to znamená, že implementace není příliš ucelená a obsahuje spoustu pozůstatků vývoje, které již nejsou využívány. Prvním krokem proto bylo odstranění těchto pozůstatků, ucelení a celkové zpřehlednění implementace modelu. Model je poměrně komplexní. Náhled uceleného modelu v softwarovém prostředí AnyLogicu můžeme vidět na Obr. 11.

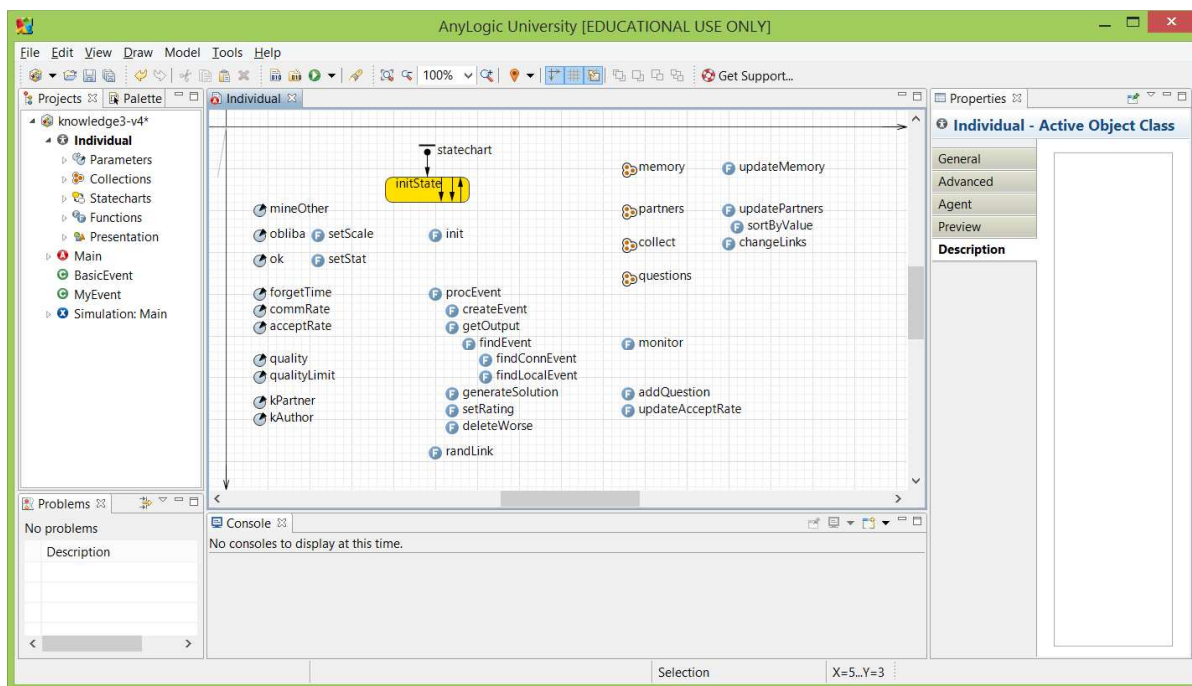
Jak je patrné, model obsahuje několik částí, a to agenta (třídu) Main, jenž zahrnuje nastavení a parametry modelu a také interpretaci výstupů (grafy), agenta Individual, který představuje našeho jedince v síti, a tím pádem nejdůležitější článek modelu, experiment simulace umožňující modifikaci počátečních vlastností modelu, a také dvě samostatné java třídy, které jsou v modelu použity pro popis událostí. Na Obr. 12 a Obr. 13 můžeme vidět náhled hlavního agenta Main a agenta Individual.



Obr. 11: Model v softwarovém prostředí AnyLogic 6.9.0



Obr. 12: Main – hlavní část modelu



Obr. 13: Agent Individual

Pro hladké pochopení a orientaci v modelu byl vytvořen velice podrobný popis dokumentující všechny tyto části a vůbec vše, co model obsahuje. Dokumentaci je možné nalézt na přiloženém CD.

Dalším krokem je transformace modelu do nové verze, jelikož AnyLogic 6.9.0 již není aktuální verzí softwaru a stávající implementace modelu není v aktuální verzi, konkrétně 7.2.0, funkční. Nefunkčnost způsobuje problém s dostupností některých vnitřních funkcí AnyLogicu, které se v nové verzi bohužel jmenují nebo volají jinak než ve staré verzi. Volání těchto funkcí bylo tedy nutné patřičným způsobem upravit.

Veškeré informace o vlastnostech, download jednotlivých verzí, případové studie, výuková videa, tutoriály a vůbec vše o simulačním softwaru AnyLogic můžeme nalézt v [16].

5.2 Model jako komplexní síť

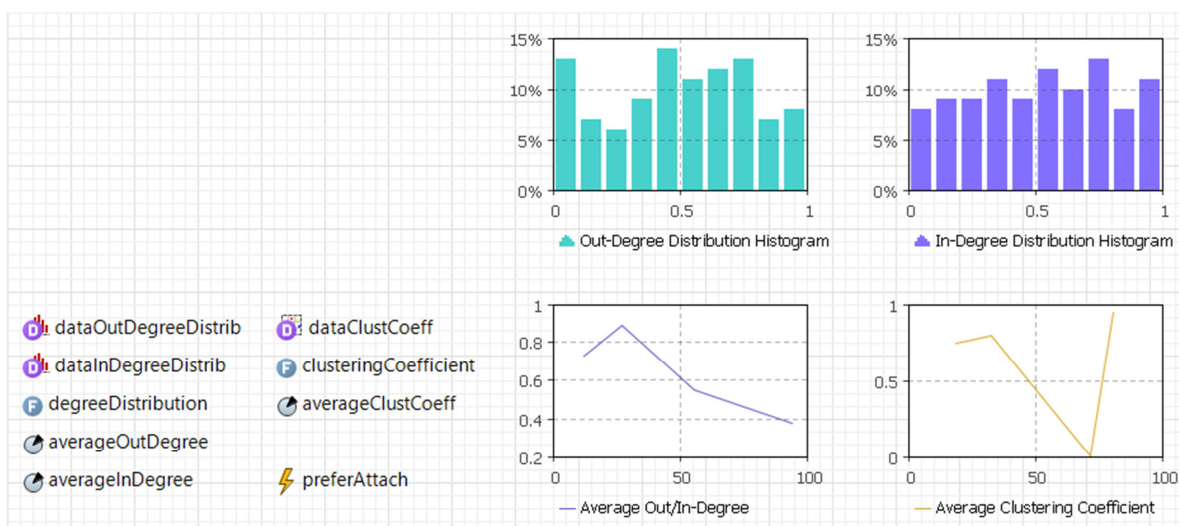
Do dnešní doby nebylo testováno, zda síť, kterou model generuje, splňuje typické vlastnosti komplexní sítě jako je například distribuce stupňů nebo shlukování.

Jak již bylo uvedeno, degree distribuce je pravděpodobnost, že náhodně vybraný uzel má určitý stupeň. To samozřejmě souvisí se stupni všech jednotlivých uzlů v síti. My se na toto nyní zaměříme v rámci našeho modelu.

Hned na začátku je potřeba si uvědomit, že vazby mezi agenty v modelu jsou orientované, čili i sám graf sítě je orientovaný. Každý agent uchovává svůj seznam partnerů, kterých se v případě vystavení události dotazuje. To ovšem neznamená, že sám agent je pro všechny své partnery také partnerem. Každý agent tedy má dva stupně - *outdegree*, znázorňující počet odchozích vazeb neboli velikost seznamu partnerů, a *indegree*, znázorňující počet příchozích vazeb neboli pro kolik agentů v síti je on sám partnerem.

Stejně tak se zaměříme na chování modelu ve vztahu k hypotéze o shlukování v síti. Opět je potřeba si uvědomit, že síť je orientovaná, to znamená, mezi každou dvojicí agentů mohou existovat dvě opačné vazby, základní poučky a přístupy je tedy nutné modifikovat.

Za tímto účelem byly do modelu doplněny různé typy grafů a vše potřebné pro jejich vykreslování. Pro představu se můžeme podívat na Obr. 14. Vidíme, že byla doplněna sada grafů, datových podkladů pro tyto grafy a parametry a funkce starající se o naplňování těchto datových podkladů.

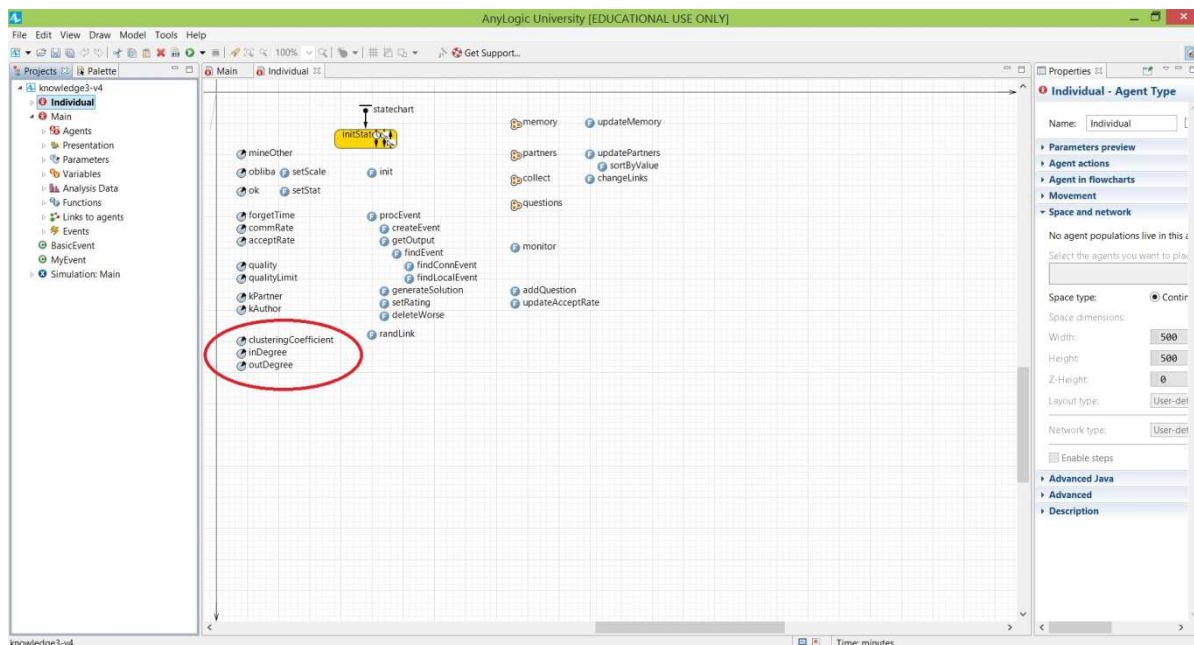


Obr. 14: Rozšíření modelu o další funkcionalitu

Budeme zjišťovat tyto hodnoty a zobrazovat je v histogramech, čímž získáme představu o rozdělení stupňů v síti. Do modelu byla rovněž implementována metrika zkoumající závislost kvality znalostí agentů na jejich stupních. Výzkumnou hypotézou zde je, že kvalita agenta významně ovlivňuje zejména jeho indegree.

Měřit budeme i shlukování u jednotlivých agentů a do modelu byl doplněn i graf znázorňující průměrný koeficient shlukování celé sítě. Stejně jako v předchozím případě budeme zkoumat i závislost kvality znalostí agentů na shlukování.

Na Obr. 15 se můžeme ještě podívat na rozšíření třídy agenta Individual o další parametry, které doposud neobsahoval. Jedná se o parametry zachycující stupně agenta (jak už víme, máme dva) a parametr znázorňující jeho koeficient shlukování.



Obr. 15: Rozšíření agenta Individual

Takovýchto kosmetických úprav modelu v podobě přidání grafu za účelem sledování vlastností nebo i vývoje těchto vlastností v průběhu času jakož i doplnění dalších parametrů, pomocných proměnných nebo celých funkcí bude v rámci práce provedena celá řada.

5.3 Chování agenta

Jak již bylo uvedeno, model byl rozvinut ve dvou základních směrech. Prvním z těchto směrů bylo zdokonalení vnitřních mechanismů chování agenta, zejména v oblasti komunikace s ostatními partnery v síti. Pojďme se nyní podívat, jakým způsobem byl původní model v tomto směru upraven. Všechny zásahy si postupně osvětlíme.

5.3.1 Rozhodování agenta

Na základě experimentů s původním modelem bylo upraveno vnitřní chování agenta ve fázi hledání co nejlepší reakce na vstupní situaci. Výchozí model preferoval využití znalostí ze sociální sítě, přesnější je však využití již v původním modelu definovaného parametru *commRate*, který charakterizuje ochotu či schopnost konkrétního agenta navazovat komunikační vazby s partnery.

Je-li v původním modelu agent vystaven situaci, kterou musí řešit, první co udělá je, že se zeptá na řešení všech svých partnerů, ta doplní řešeními nalezenými ve své vlastní paměti a použije nejlepší řešení z této množiny. Pouze v případě, kdy mu ani jeden z partnerů neposkytne řešení dané situace a nenalezne-li žádné řešení ani ve své paměti, generuje agent na základě své inteligence své vlastní řešení.

Tento způsob neodpovídá příliš realitě, protože i když by agent s vysokou inteligencí byl schopen vygenerovat mnohem lepší řešení, než mu poskytují jeho partneři, přijímá jejich horší řešení. Agent by měl mít vždy možnost uplatnit svou inteligenci.

Způsob rozhodování agenta byl proto modifikován tak, že generování vlastního řešení je první akcí v případě vystavení události. V druhém kroku se agent na základě své ochoty komunikovat (která je v modelu implementována), rozhodne, zda bude kontaktovat své partnery či nikoliv. V případě malé ochoty ke komunikaci použije buď nejlepší řešení ze své paměti anebo právě vygenerované řešení, pakliže to dosahuje ještě lepší kvality. Na tomto místě je ještě potřeba si uvědomit, že při přijímání řešení hraje roli také verifikace řešení, to znamená verifikované řešení má přednost před jakýmkoli neverifikovaným. V opačném případě množinu možných řešení doplní ještě o řešení získané od svých partnerů a použije opět to nejlepší z nich. Tímto podpoříme agenty ve využívání vlastní inteligence, která mohla být často zastíněna použitím řešení s nízkou kvalitou od partnerů.

Lze namítnout, že i tímto způsobem můžeme zhoršit rozhodování agentů, protože agenti, a to převážně agenti s nízkou inteligencí, kteří nebudou kontaktovat své partnery, často přijmou mnohem horší řešení než by mohli získat aktivní komunikací. To je samozřejmě pravda, ovšem na tomto místě je potřeba si uvědomit, že naším cílem není vytvořit model, ve kterém budou agenti v každé situaci přijímat to nejlepší možné řešení, ale model, který bude co nejlépe simulovat reálný svět, kde je popsána situace zcela běžná. Je tedy více respektována realita života a rozmanitost osobností agentů ve formě vloženého stochastického prvku.

5.3.2 Úprava ochoty odpovídat a snižování ratingu

Další provedenou úpravou byl systém pro průběžnou aktualizaci hodnoty *acceptRate*, která byla také zavedena již v původním modelu a charakterizuje ochotu agenta odpovídat na jemu pokládané dotazy na řešení od ostatních agentů. Tuto hodnotu lze chápat jako maximální pravděpodobnost, se kterou agent na danou otázku odpoví. Maximální proto, že v reálném světě je běžné, že u jedinců, kteří budou extrémně zatěžováni dotazy, bude ochota k

poskytování odpovědí klesat s rostoucím počtem dotazů. Tento jev byl implementován do modelu formou průběžné aktualizace hodnoty dle vzorce (14).

$$a_{new} = \min \left(a_{orig}, a_{old} \left(1 + k_a \frac{n_{opt} - n}{n_{opt}} \right) \right) \quad (14)$$

Ve vzorci je a_{new} nová hodnota parametru *acceptRate*, k_a koeficient vlivu odchylky od ideálně předpokládaného počtu otázek v jednom kroku n_{opt} a n počet položených otázek. Symbol a_{orig} označuje hodnotu *acceptRate* nastavenou na začátku simulace jako osobnostní charakteristiku agenta a a_{old} aktuální hodnotu. Parametr se tedy pohybuje v intervalu $\langle 0; a_{orig} \rangle$ namísto původního intervalu $\langle 0; 1 \rangle$. Průměrná ochota agentů reagovat na pokládané dotazy by tedy měla klesnout, protože agenti, kterým bude hodnota *acceptRate* respektive a_{orig} na počátku simulace nastavena například na hodnotu 0,1 se nad tuto hodnotu nikdy nedostanou, a to ani v případě nízkého počtu dotazů na ně. To znamená, že neochota těchto agentů není způsobena velkým množstvím dotazů (s těmi se může maximálně ještě více snížit), ale zkrátka tím, že neochota někomu na něco odpovídat je jejich charakteristická vlastnost.

Komunikace agentů se rovněž týká úprava spočívající ve snížení ratingu partnera, který byl agentem dotázán, avšak neodpověděl. To souvisí i s výše uvedenou úpravou *acceptRatu* a snížením ochoty odpovídat. Důsledkem tohoto kroku je postupné posouvání partnera níže v seznamu partnerů a v případě překročení maximální délky tohoto seznamu jeho vyřazení. Samozřejmě pouze v případě, že dotazovaný partner opakovaně odmítá odpovídat. Za každé nezodpovězení dotazu je rating snížen o 10 procent.

5.3.3 Partnerství jako efekt zodpovězení dotazu

Dlouho řešený problém, který se vyskytoval v rámci této práce, byl ten, že agenti vůbec nevyužívali celkovou kapacitu pro svůj seznam partnerů. Jak již víme, velikost tohoto seznamu může být modelem omezena a v našich experimentech byla povětšinou omezena hodnotou 20, avšak žádný agent tolika partnery nedisponoval a ono omezení tedy, dalo by se říci, postrádalo smysl. Další uvažovaný problém, který by mohl v této situaci vznikat, je omezení šíření informací v síti právě z důvodu nedostatku vazeb, kterými agenti disponují.

Poslední provedenou úpravou a řešením výše popsaného problému bylo zahrnutí agenta pokládajícího dotaz, který byl zodpovězen, do seznamu partnerů agenta dotazovaného, s tím že rating tohoto partnera je nastaven na minimální hodnotu. Tento

postup dobře simuluje skutečnost, kdy pokud jsme dotázáni, obvykle si uvědomujeme, kdo se ptal a jsme schopni kontakt na něj využít v další komunikaci. Je potřeba zdůraznit, že k partnerství dojde, jak bylo zmíněno, pouze v případě, že dotazovaný agent je ochoten dotaz zodpovědět. V případě neochoty z důvodu nízké hodnoty parametru *acceptRate* (viz výše) dotazující se agent do seznamu partnerů dotazovaného zahrnut není. Vše je opět děláno s cílem co nejlépe popsat chování osob v reálném světě.

5.4 Eliminace předpokladu uzavřeného světa

Dostáváme se k druhé oblasti zdokonalení modelu. Jak již víme, tou byla úprava celého modelu, který byl konstruován s předpokladem tzv. uzavřeného světa, kdy nebylo možné měnit počet situací, jimž byli agenti vystavováni, ani počet těchto agentů. V následujících částech práce se na jednotlivé zásahy podíváme. Opět si vše postupně osvětlíme.

5.4.1 Rostoucí síť a upřednostňované připojování

Předpoklad uzavřeného světa byl jedním ze základních rysů původního modelu. V navržené úpravě je tento předpoklad eliminován. Začneme s eliminací neměnného počtu uzlů. Návrh této úpravy měl svůj vývoj a sestával z několika dílčích na sebe navazujících úprav. Pojdme se nyní podrobně podívat, jakým způsobem probíhalo zavedení rostoucí sítě.

Upřednostňované připojování aplikované na celou síť

Původní model pracoval s pevně definovaným počtem uzlů, který se v celém průběhu simulace neměnil. První myšlenka byla taková, že tento počet budeme každý simulační krok inkrementovat o jeden uzel neboli o jednoho agenta, to znamená, například v časovém kroku $t=100$ bude populace obsahovat 200 agentů, v kroku $t=500$ potom 600 agentů atd. Můžeme ještě podotknout, že podobně jako velké on-line sociální sítě i my předpokládáme, že z těchto systémů jedinci v podstatě neodcházejí, zato však významnou měrou přibývají. Otázkou ale zůstává, jakým způsobem nově příchozího agenta do stávající sítě zapojit. Z teoretické části víme, že existuje předpoklad upřednostňovaného připojování (angl. Preferential Attachment), který říká, že nově příchozí jedinec do sítě vytvoří vazbu na m uzlů s pravděpodobností úměrnou jejich stupni. Uděláno to bylo přesně tak.

Do modelu byla implementována událost, která v každém simulačním kroku přidá do sítě jednoho agenta. Nově příchozí agent vytvoří vazbu pouze s jedním uzlem, v našem případě tedy $m=1$, respektive přijme ze stávající sítě za svého partnera jednoho agenta s pravděpodobností úměrnou součtu odchozích a příchozích vazeb (outdegree + indegree).

To znamená, že čím větší je počet jedinců, se kterými agent komunikuje (bez ohledu na směr komunikace), tím větší je pravděpodobnost, že se stane partnerem nově příchozího agenta.

Upřednostňované připojování aplikované lokálně

V průběhu naší práce jsme se ovšem nad touto věcí pokoušeli zamyslet více do hloubky a naší snahou bylo opět co nejvíce korelovat s děním v reálném světě. Další modifikací, respektive zdokonalením upřednostňovaného připojování je, že ho budeme aplikovat pouze lokálně na nejbližší okolí nově příchozího agenta. Agenti se v modelu vyskytují v určitém prostředí a každý má svou vlastní pozici. Můžeme si to představit jako bydliště každého člověka v reálném světě. Nově příchozímu agentu je přidělena náhodná pozice v tomto prostoru a agent, ke kterému je připojen, je vybrán stejným upřednostňovaným způsobem, ovšem již pouze z jeho nejbližšího okolí (řekněme 10 nejbližších agentů).

Tímto způsobem bychom se se svým modelem mohli opět více přiblížit reálnému světu, protože uvážíme-li to, že se někdo například přestěhuje do jiného města, čímž vlastně přichází do nové sociální sítě, s kým se asi nejprve seznámí? S největší pravděpodobností s někým z blízkého okolí, ať už se sousedem nebo někým ze stejné či sousední ulice. Ne že by to nebylo možné, ale je velmi málo pravděpodobné, že první s kým přijde do styku, bude někdo z druhého konce města. Událost navyšující počet agentů byla proto patřičným způsobem upravena.

I nyní nově příchozí agent preferenčně přijme za svého partnera jednoho agenta, tedy $m=1$, ovšem ten je vybírán již pouze z jeho nejbližšího okolí. V našem případě toto okolí tvoří 10 nejbližších agentů v rámci 2D vizualizace. Je tedy určitým způsobem respektována lokalita, ve které se agent vyskytuje.

Prosté připojení na nejbližší sousedy

V rámci těchto úprav byla tato skutečnost zkoumána ještě z jiného pohledu. Uvažovali jsme nad samotným smyslem preferenčního připojování. Je to skutečně způsob, jakým reálné sociální sítě vznikají? Jak již velmi dobře víme, preferenční připojování velice dobře vysvětluje mocinné rozdělení stupňů v síti. To ale nemusí nutně znamenat, že vývoj vazeb v reálných sociálních sítích je doopravdy tímto řízen. Zamysleme se nad tím, jakým způsobem lidé navazují kontakty. Je zde zásadním vlivem opravdu počet již existujících vazeb? Určitý vliv to mít jistě bude. Je mnohem pravděpodobnější, že člověk naváže kontakt s někým, s kým komunikuje již mnoho lidí než s někým, kdo mezi lidmi není moc známý v tom smyslu, že disponuje malým množstvím kontaktů. Ale zcela jistě toto nebude jediným

faktorem, který ovlivňuje dynamiku vzniku vazeb. Už jenom z toho důvodu, že v momentě příchodu mnohdy ani neznáme počet vazeb, kterým jedinci disponují. Oněch faktorů bude existovat daleko větší množství. Zamyslíme-li se, bude to jistě sympatie, krása, chytrost, vtipnost, pochopení a mnoho dalších. Tím se ale dostáváme do extrémní složitosti reálného světa, které se v tuto chvíli vyhneme.

Co chceme říci je to, že základ dynamiky vzniku sociálních sítí, která vede k mocinnému rozdělení stupňů, nemusí být nutně tvořen upřednostňovaným připojováním popisovaným výše, respektive lze ukázat, že modelovaná síť se bude takto vyvíjet i bez jeho přímé implementace.

Poslední úpravou růstu sítě je tedy to, že v libovolném okamžiku simulace, v našem případě v každém simulačním kroku, je možné vložit nového jedince a seznam jeho partnerů inicializovat na m jeho nejbližších sousedů dle 2D vizualizace. Zde vycházíme z názoru, že agent či jedinec vložený do prostředí sociální sítě se bude v prvním kroku snažit zejména navázat vazby a vztahy s jemu blízkými či známými jedinci, které se budou teprve následně vyvíjet tak, aby byla pro tohoto jedince sociální síť co nejpříznivější, tedy aby byl schopen z ní získávat kvalitní řešení situací, kterým je vystavován (viz preferenční připojování zmíněné výše). Vliv počtu stávajících vazeb jednotlivých uzlů byl tedy eliminován a budeme sledovat dynamiku vazeb řízenou prostým napojením nově příchozích jedinců na jim nejbližší jedince.

Finální úprava události je tedy taková, že nově přicházející agenti již nevolí za své partnery agenty s nejvyšším stupněm, ale jednoduše navážou kontakt s m , v našem případě $m=5$, nejbližšími z nich, což mimochodem koresponduje s velikostí první vrstvy ego modelu, který byl popisován v teoretické části.

5.4.2 Flexibilní počet situací

Dostáváme se k druhé úpravě v rámci eliminace předpokladu uzavřeného světa a tou je eliminace fixního počtu situací.

Zavedení flexibilního počtu situací, na které agent musí reagovat, bylo vyvoláno snahou přesněji modelovat realitu - i my jsme vystavováni podnětům či situacím, které jsme do tohoto okamžiku nezažili a snažíme se s nimi vypořádat. Nové situace tedy mohou být do modelu zaváděny i v průběhu simulačního procesu.

Nicméně pro přiblížení se realitě bylo nutné provést ještě zásadní úpravu v oblasti generování vlastních řešení agentem. Tento proces dosud předpokládal nastavení tzv. kvality agenta, která v podstatě vyjadřovala schopnost agenta generovat kvalitní řešení situací danou

úrovni jeho inteligence a znalostmi. Vyjdeme-li však z předpokladu, že řešené situace nepocházejí ze společné (znalostní) domény, je velmi reálné, že jedinec nebude schopen na všechny tyto události reagovat stejně kvalitně neboli, mohli bychom říci, v každé situaci bude jinak inteligentní.

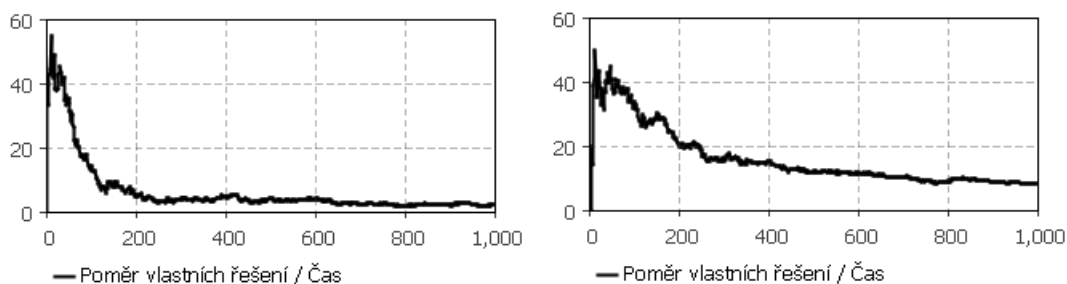
Provedená úprava modelu spočívá v tom, že každý agent nebude disponovat pouze jedním parametrem kvality, ale kvalita generovaných řešení bude pro každou událost jiná. Nebo, jinými slovy, agenti budou v řešení určitých situací inteligentnější než v řešení jiných. To se přibližuje reálnému světu, ve kterém každý umí něco, respektive něco lépe a něco hůře. Z toho důvodu je nyní kvalita či úroveň znalostí agenta nastavována pro každou situaci samostatně.

6 Výsledky experimentů

V této kapitole jsou představeny výsledky experimentů provedených s modifikovaným modelem popsaným v předchozí kapitole. Všechny experimenty, pokud není uvedeno jinak, byly provedeny s nastavením sítě na 1000 simulačních kroků a s maximálním počtem partnerů agenta nastaveným na 20. Počet počátečních vazeb agenta $m=5$. Rostoucí sítí je v popisu myšlena síť o počátečním počtu uzlů 1 rostoucí s každým simulačním krokem o jednoho agenta.

6.1 Rozhodování agenta

V prvním experimentu se podíváme na to, jak modifikace způsobu rozhodování ovlivnila agenty. V původním modelu byl implementován parametr agenta, který udává poměr svých vlastních řešení ke všem řešením v paměti agenta. Každý objekt řešení v sobě totiž obsahuje informaci o tom, kdo dané řešení vygeneroval čili o autorovi řešení. Toho bylo využito k porovnání změn. Na Obr. 16 můžeme vidět průměrný vývoj poměru vlastních řešení v celé síti před a po úpravě modelu.



Obr. 16: Vývoj poměru vlastních řešení před a po úpravě způsobu rozhodování agentů

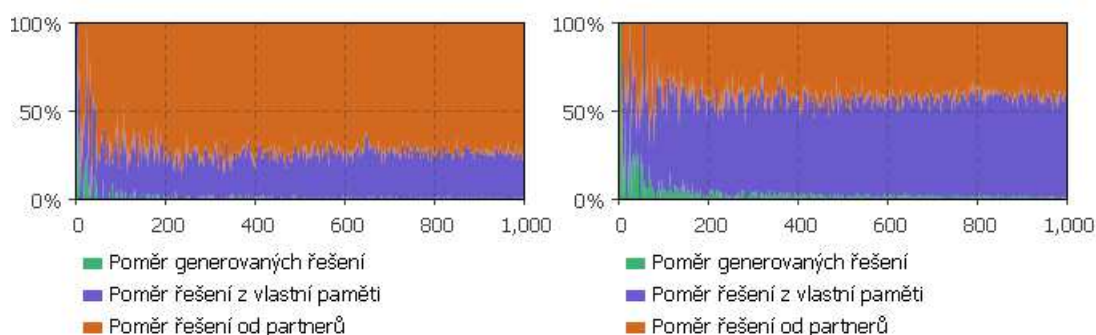
Je zřejmé, že na počátku simulace, kdy agentům často nezbyvá nic jiného než generovat vlastní řešení, je poměr vlastních řešení v paměti vysoký, nicméně velmi brzy začne klesat.

V levém grafu je zřejmý prudký pokles, kdy hodnota zhruba v kroku 200 klesne téměř k nule a mírně klesá i ve zbytku simulační doby, to znamená, jakmile je síť zaplavena nějakými řešeními, agenti slepě přejímají ta nejlepší z nich a nemají již téměř žádnou snahu, s konstrukcí původního modelu ani mít nemohou, generovat ještě lepší řešení. Podotkněme, že experiment byl prováděn pouze s jednou situací vyžadující řešení.

V pravém grafu je znázorněna ta samá situace, ovšem zde již bylo pracováno s upraveným způsobem rozhodování agentů, kde byl důraz na využití jejich vlastní inteligence. Opět vidíme situaci, kdy na počátku agenti často generují svá vlastní řešení, ovšem již nedochází k tak prudkému poklesu jako v předchozím případě. To je důkaz toho, že nyní agenti mnohem více využívají svou vlastní inteligenci. Ovšem i zde můžeme vidět klesající charakter hodnoty po celou dobu simulace. To můžeme vysvětlit jednoduše tím, a je to také důkaz toho, že v obou případech v našich sítích dochází k šíření informací, kdy se agenti snaží získat ze sítě co nejvíce. To znamená, postupem času přijmou řešení od těch nejinteligentnějších agentů v síti, kterých nemusí být mnoho, a svá vlastní řešení zapomínají. Na konci simulace tedy paměť většiny agentů obsahuje řešení, které vygeneroval jiný (inteligentnější) agent.

V rámci tohoto experimentu byl do modelu implementován ještě další detailnější graf demonstrující vliv úpravy rozhodování agentů. Je-li agent vystaven události, má, jak již jistě víme, tři možné způsoby, jak přijmout řešení, a to ze své vlastní paměti, komunikací neboli od partnerů anebo generuje řešení nové.

Na Obr. 17 můžeme vidět grafy, které znázorňují, jakým způsobem agenti nejčastěji přijímají řešení situací. Opět máme pro porovnání grafy dva. Graf vlevo znázorňuje chování modelu před úpravou a vpravo po úpravě.



Obr. 17: Poměry zdrojů řešení před a po úpravě způsobu rozhodování agentů

Tyto grafy jsou postaveny trochu jinak než v předchozím případě, kde jsme analyzovali obsah paměti agentů. Zde se zaměřujeme, na přírůstek paměti neboli jednoduše v každém časovém kroku zaznamenáváme to, jakým způsobem agenti přijímali řešení.

Z levého grafu je opět patrné, že na počátku simulace mají agenti, protože jim ani nic jiného nezbyvá, snahu generovat řešení nebo využívat svou vlastní paměť. Tato snaha není ale nijak velká, jakmile je v síti k dispozici dostatek řešení, pracují agenti již pouze s nimi a

generování dalších prakticky úplně vymizí. To je v souladu s předchozí analýzou. Všimnout si můžeme i jevu, že po zaplavení sítě řešeními využívají agenti zhruba ve 25 % případů svou vlastní paměť a v 75 % přijímají řešení od svých partnerů. To může poukazovat na to, že existuje 75 % pravděpodobnost, že agent získá lepší řešení prostřednictvím komunikace než ze své vlastní paměti.

Po provedené úpravě se chování modelu poněkud změnilo. Výsledek vidíme opět v pravém grafu. Na začátku simulace agenti vykazují mnohem větší snahu generovat svá vlastní řešení. To je opět shodné s předchozí analýzou. Nicméně generování vlastních řešení upadá a postupem času opět téměř úplně vymizí. Tento fakt je podporován také tím, a na tomto místě je potřeba připomenout, že při přijímání řešení hraje roli také verifikace a vzhledem k tomu, že každé nově vygenerované řešení je neverifikované, je toto překonáno jiným již verifikovaným řešením.

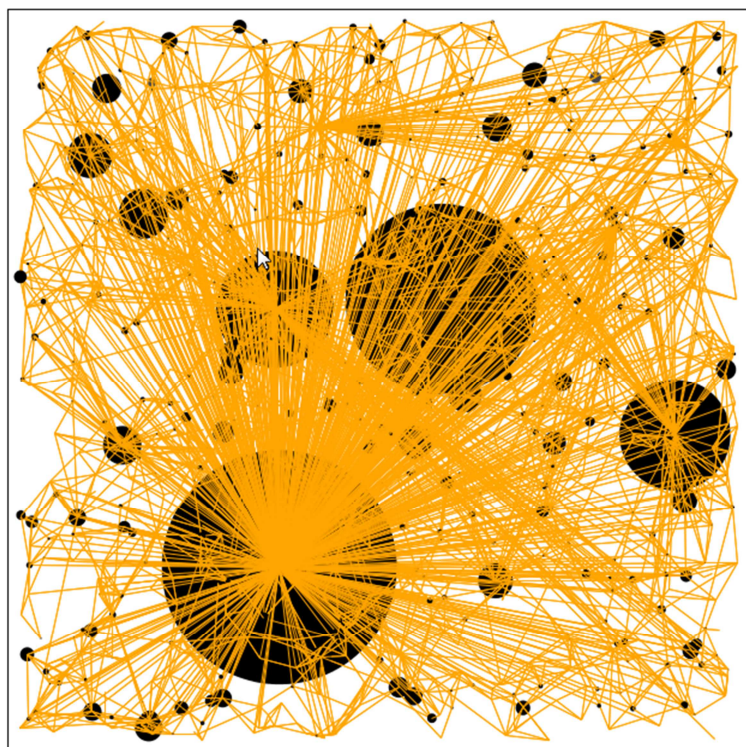
Změna, kterou nelze přehlédnout, je, že agenti nyní mnohem více využívají svou vlastní paměť. To je způsobeno zohledněním jejich ochoty ke komunikaci. Nízká ochota mnoha agentů způsobuje, že agenti při řešení situací vůbec nekontaktují své partnery a využívají pouze svou vlastní paměť. Můžeme ještě podotknout, že agenti mohou ze své paměti klidně přijmout i řešení, jehož autorem je někdo jiný.

6.2 Úprava ochoty odpovídat a snižování ratingu

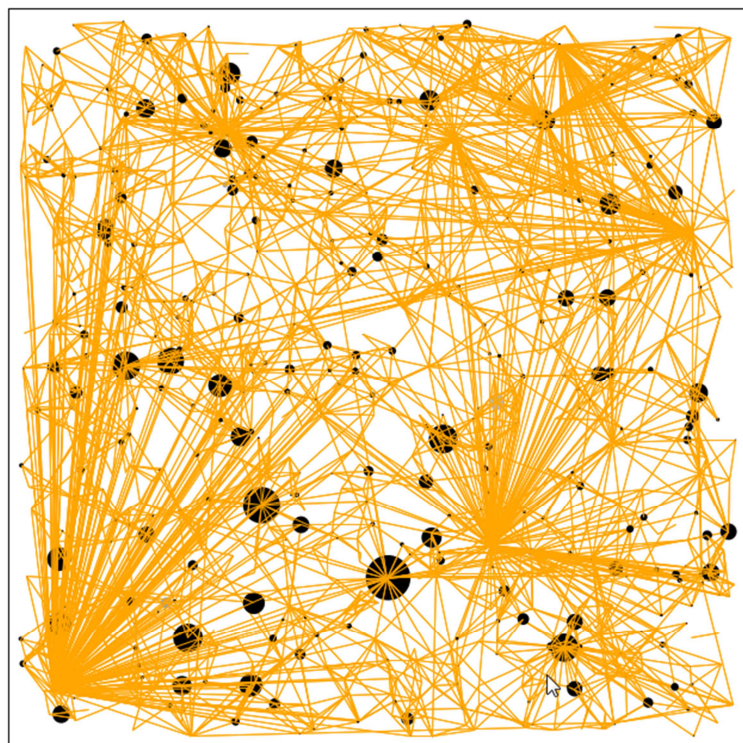
V dalším experimentu se podíváme na výrazné změny ve struktuře sítě, které přinesla úprava chování agentů při velkém množství dotazů na ně. Strukturu sítě původního modelu můžeme vidět na Obr. 18.

Struktura sítě v tomto případě odpovídá síti s výraznými jedinci se značnou kapacitou odpovídat na dotazy. Velikost uzlu odpovídá jeho oblíbenosti v síti, jejíž hodnota je dána součtem ratingů daného agenta u všech ostatních agentů v síti, respektive u těch, pro které je daný agent partnerem. Takováto síť také velmi dobře ilustruje preferenční připojování, ale nereflektuje stav, kdy agent může být dotazy, které jsou mu kladeny, přetížen. Hodnota ratingu partnera se pohybuje v intervalu $<0;1>$ podle toho, jak kvalitní řešení dodává, to znamená, takto vysoké oblíbenosti může agent dosáhnout pouze v případě, že je partnerem pro velké množství agentů v síti, kteří ho zaplavují dotazy, a jeho oblíbenost tím pádem neustále roste.

Po provedené úpravě chování zohledňující klesající ochotu odpovídat na otázky v případě jejich velkého počtu a s tím spojeného snižování ratingu se struktura jinak stejně modelované sítě výrazně změnila. Změny je možné shlédnout na Obr. 19.



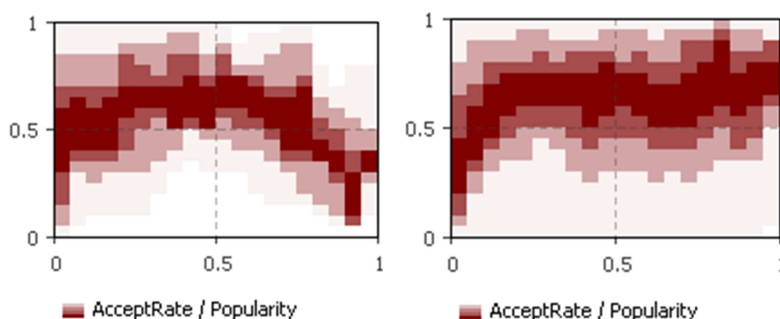
Obr. 18: Struktura sítě dle původního modelu



Obr. 19: Struktura sítě dle inovovaného modelu

Je patrné, že upravený model lépe odpovídá síti víceméně podobných jedinců, kde implementovaný mechanismus nedovoluje extrémní zatížení jedinců dotazy – respektive dovoluje, ale většina z nich nebude zodpovězena – a tedy ani jejich extrémní oblibu z důvodu snižování ratingu za nezodpovězené dotazy.

Funkčnost tohoto mechanismu ukazuje i 2D histogram závislosti ochoty agenta odpovídat (*acceptRate*) na jeho oblibě (*popularity*) ve fázi vývoje sítě, kdy agenti ve velké míře získávají řešení komunikací a hodnota *acceptRate* pro oblíbené agenty je redukována, a v ustálené fázi, kdy již disponují řešeními z vlastní paměti, komunikace není tak intenzivní a nemusí být regulována (Obr. 20). Počty agentů v jednotlivých oblastech jsou znázorněny sytostí zbarvení (sytější barva – více agentů v daném poli).



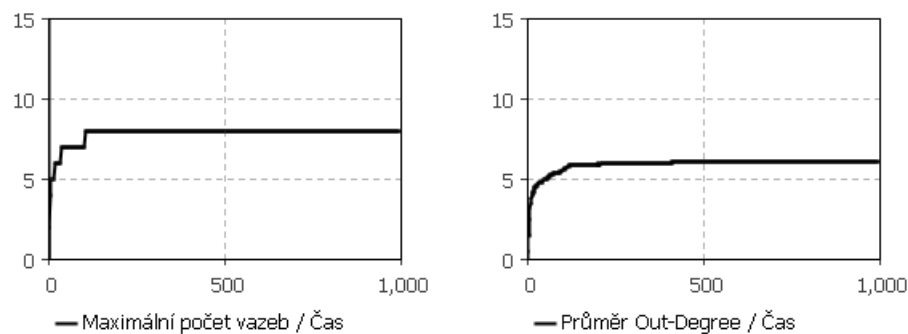
Obr. 20: 2D Histogram závislosti ochoty agenta odpovídat na jeho oblibě

6.3 Partnerství jako efekt zodpovězení dotazu

Pojďme si ukázat, jaký vliv měla poslední úprava v rámci vnitřních mechanismů chování agentů. Řešili jsme problém s tím, že agenti v síti disponují malým množstvím vazeb, respektive plně nevyužívají již tak omezenou kapacitu seznamu partnerů.

Vývoj počtu partnerů agentů před provedením modifikace modelu můžeme vidět na Obr. 21. K dispozici máme dva grafy, první zachycuje maximální velikost seznamu partnerů v síti v čase a druhý průměrnou velikost neboli průměrný počet odchozích vazeb v celé síti v čase.

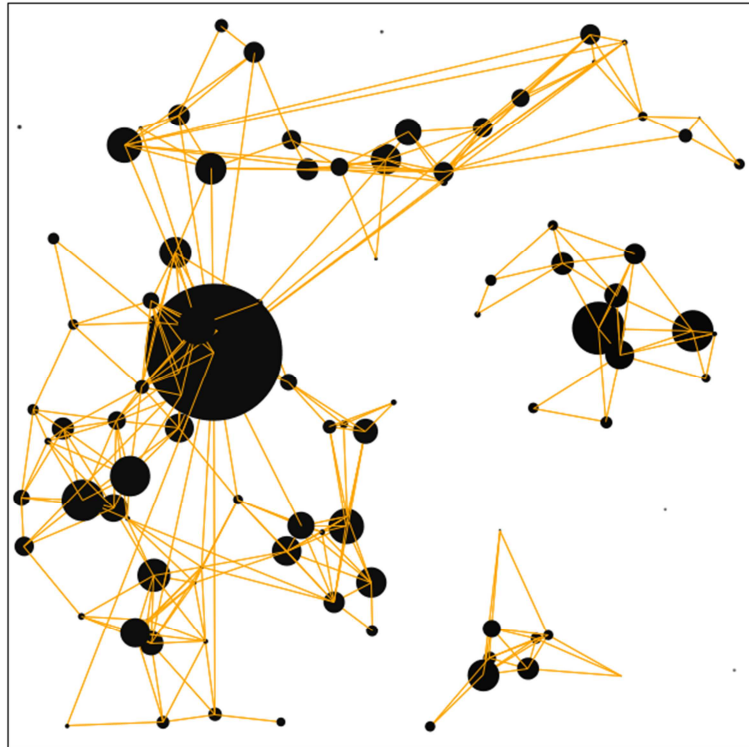
Z obrázku je patrné, jak málo vazeb agenti vytvářeli. V levém grafu lze vidět, že maximální velikost seznamu partnerů se krátce po spuštění simulace vyšplhala na hodnotu 8, ovšem po celý zbytek simulačního času se již nezvýšila. Pravý graf nám ukazuje, že průměrná velikost seznamu partnerů roste, podobně jako ta maximální, na začátku simulace, avšak krátce po začátku se ustálí zhruba kolem hodnoty 6 a po zbytek doby se již také téměř nemění. Jako kdyby něco bránilo v získávání dalších vazeb.



Obr. 21: Vývoj velikosti seznamů partnerů před úpravou modelu

Zamysleme se nejprve nad tím, jakým způsobem v původním modelu agenti získávají své partnery. Partnerství vzniká pouze jedním jediným způsobem a to prostřednictvím původců neboli autorů řešení. Jinými slovy, přijme-li agent řešení od svého partnera, který v ten moment hraje roli dodavatele řešení, stává se agentovým partnerem i autor tohoto řešení, samozřejmě pakliže není totožný s dodavatelem a nevyskytuje-li se již v seznamu partnerů agenta.

Ovšem aby agent mohl od někoho přijímat řešení, musí ho již obsahovat ve svém seznamu partnerů. V případě, že bude existovat agent, který nemá ani jednoho partnera, nemá po celou dobu simulace ani možnost žádného získat. To samozřejmě závisí na struktuře sítě na počátku simulace, která byla v původním modelu taková, že konstantní počet v prostoru náhodně rozmístěných agentů byl spojen vazbami na základě vzdálenosti, to znamená, že se mohli vyskytnout agenti nedisponující žádnou vazbou, a tudíž odsouzeni k tomu, že jim při řešení situací po celou dobu simulace nikdo neporadí. Stejně tak mohlo dojít k vytvoření několika vzájemně nepropojených shluků – komponent, mezi nimiž v takto postaveném modelu nikdy žádná komunikace probíhat nebude. Náhled takovéto struktury sítě s neměnným počtem uzlů původního modelu je znázorněn na Obr. 22, kde si lze všimnout několika izolovaných uzlů a tří vzájemně nepropojených komponent sítě.



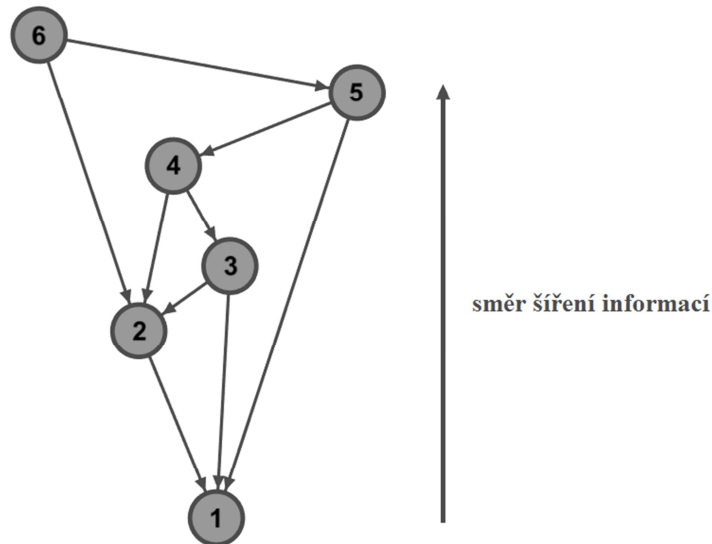
Obr. 22: Struktura statické sítě původního modelu

Podotkněme ale, že náš experiment byl prováděn již s rostoucí sítí, ve které se výše popsaný problém již nevykazuje, protože se v průběhu simulace objevují nové uzly. Nicméně ihned narážíme na další problém. I v rostoucí sítí, které se budeme detailně věnovat v dalších experimentech, totiž existuje silné omezení partnerství některých uzlů kvůli jednosměrnosti nebo, chceme-li, orientovanosti vazeb v síti.

Problém orientovanosti spočívá v tom, že nově příchozí uzel je napojen, bez ohledu na to, jakým způsobem, na m , řekněme např. dva, stávajících uzlů. Co to znamená? Pro lepší představu a pochopení se podívejme na Obr. 23.

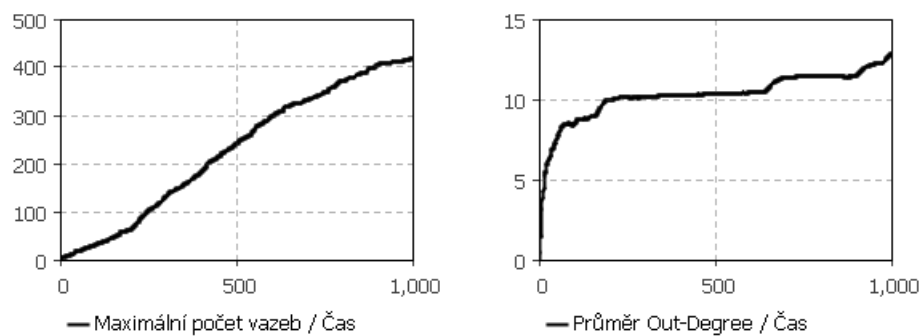
Vidíme rostoucí síť o šesti uzlech, respektive agentech, kteří jsou očíslovány podle toho, jak do sítě přicházeli. Každý příchozí agent přijal ze stávající sítě za své partnery dva agenty. Problém naznačuje směr šíření informací. Vzhledem k tomu, že partnerství vzniká, jak víme, pouze na základě autorství, nově příchozí agenti se nikdy nestanou partnery již stávajících agentů. Uvedme příklad. Agent 4 má ve svém seznamu partnerů agenta 2 a 3. Jeho partnerem se časem může stát i agent 1, a to v případě, že přijme jím vygenerované řešení (tj. ono autorství) právě z paměti agenta 2 nebo 3, pro něž oba je agent 1 partnerem. Ovšem jeho partnery se nikdy nemůžou stát agenti 5 a 6, protože jimi vygenerovaná řešení se nikdy nedostanou do paměti agentů 1, 2 ani 3. Vzniká jakási stromová struktura, ve které se informace, jak je naznačeno šipkou, šíří zdola nahoru a nově příchozí agenti jsou, podle

toho, ve které části se napojí, odsouzení využívat pouze informace od agentů, kteří se v této stromové struktuře nachází pod nimi. Tento problém byl úpravou modelu vyřešen.



Obr. 23: Problém orientovanosti vazeb

Pojďme se podívat na to, jak se změnil vývoj počtu partnerů agentů po výše popsané úpravě modelu, kdy došlo k určité eliminaci omezení způsobeného orientovaností vazeb. Výsledek můžeme vidět na Obr. 24.



Obr. 24: Vývoj velikosti seznamů partnerů po úpravě modelu

Vliv úpravy je zřejmý na první pohled a to obzvláště v levém grafu znázorňujícím maximální velikost seznamu partnerů v síti, která se vyšplhala nad hranici 400. Podotkněme, že v tomto experimentu bylo omezení velikosti seznamu partnerů samozřejmě vypnuto, respektive nastaveno na hodnotu 1000. Ovšem levý graf zachycuje maximální velikost, to

znamená, takovou velikostí seznamu partnerů disponuje pouze jeden nebo několik málo agentů vzhledem k tomu, že průměrná velikost seznamu partnerů, jak ukazuje graf vpravo, se na konci simulace pohybuje zhruba kolem hodnoty 12. Průměrně by tedy agenti kapacitu seznamu partnerů nastavenou na hodnotu 20 nevyužili ani po provedené úpravě modelu, musíme si ale uvědomit, že se jedná o průměr celé sítě a mnoho agentů v tomto případě tuto hodnotu jistě přesahuje, což může napovídat i extrémní hodnota maximální velikosti v levém grafu. Omezení velikosti seznamu partnerů by tedy již mělo smysl.

Upozorníme ještě na to, že nízký průměr velikosti seznamu partnerů může být způsoben také tím, a to je téměř jisté, že experiment byl prováděn pouze s jednou jedinou situací k řešení, to znamená, všichni agenti v síti řeší pouze jeden a ten samý problém, síť je zaplavena mnoha řešeními tohoto problému a agentům k uspokojení svých potřeb postačí navázat jen několik málo vazeb. Zvýšením počtu situací se také v našich experimentech budeme věnovat.

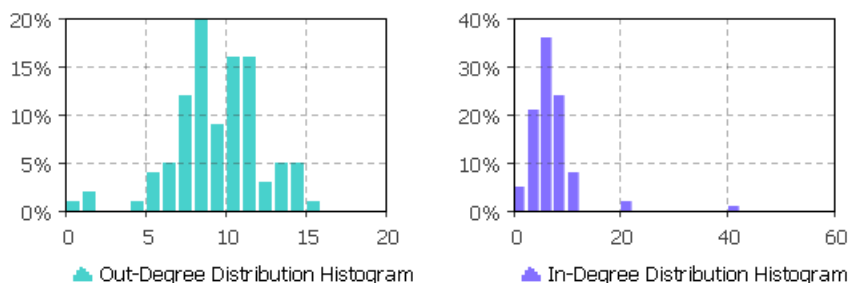
V pravém grafu si můžeme ještě všimnout náhlých vzrůstů v průměrném počtu odchozích vazeb. To může být vysvětleno tím, že se v síti objevilo velmi kvalitní řešení, kvalitnější než všechna předchozí, které vygeneroval agent s adekvátní inteligencí a které se díky úpravě modelu mohlo v síti snadno a rychle rozšířit a všichni agenti, ke kterým se toto řešení dostalo, okamžitě navázali kontakt s jeho autorem.

Zauvažujme ještě nad tak extrémní maximální velikostí seznamu partnerů v levém grafu. To může totiž souviset s náhlými nárůsty v průměru odchozích vazeb (*outdegree*), kdy mnoho agentů navazuje kontakt na autora velmi kvalitního řešení a ten tím pádem disponuje velkým množstvím příchozích vazeb (*indegree*). V případě reakce autora na všechny tyto příchozí vazby v následné komunikaci se všichni tito agenti stávají jeho partnery, což vysvětluje onu extrémní hodnotu velikosti seznamu partnerů. To je jasný příklad preferenčního připojování, kdy se většina sítě orientuje na jednoho nebo několik málo jedinců. Tím se také v následující části práce budeme zabírat.

6.4 Rostoucí síť a upřednostňované připojování

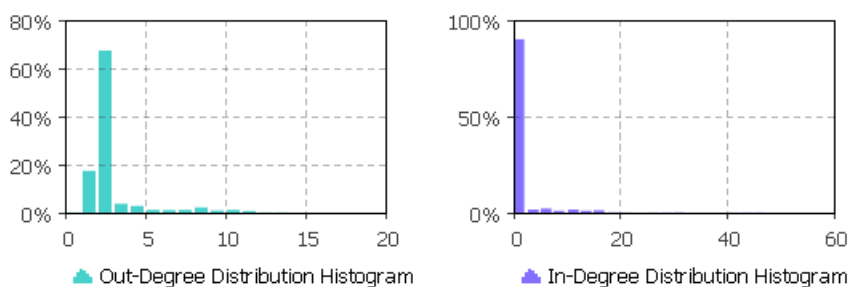
Dostáváme se k problematice rostoucí sítě. Pojďme se nyní podívat na to, jakým způsobem úpravy v tomto směru ovlivnily náš model. Stejně jako ve všech předchozích i následujících experimentech využijeme ke sledování chování modelu doimplementovaných grafů. V prvním experimentu bylo zkoumáno pravděpodobnostní rozložení hodnot stupňů uzlů sítě. Zde je potřeba si připomenout, jak bylo zmíněno v části 5.2, že vazby mezi agenty v modelu jsou orientované a můžeme tedy rozlišit stupně dva.

Na Obr. 25 můžeme vidět histogram rozdělení hodnot stupňů ve statické síti o 100 agentech vystavovaných jediné situaci v původním modelu. K dispozici máme jak distribuci pro odchozí vazby (*outdegree*), tak distribuci pro příchozí vazby (*indegree*).



Obr. 25: Degree distribuce na statické síti

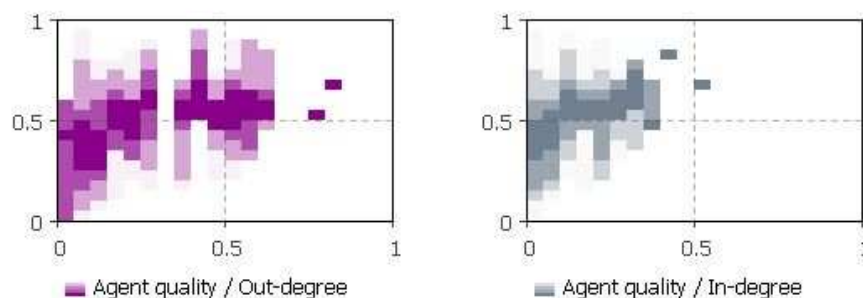
Histogramy se svým tvarem blíží Poissonovu rozdělení, což odpovídá náhodným grafům, které vykazují určité maximum, kolem kterého se pohybují stupně všech uzlů. Z teoretické části ovšem víme, že rozsáhlé sociální sítě jsou bezškálové (*scale-free*) a rozdělení stupňů v nich by mělo být zcela odlišné. Na Obr. 26 můžeme naproti tomu vidět rozdělení stupňů v rostoucí síti, jejíž parametry byly shodné s předchozí, ale byla již na upraveném modelu popsaném na začátku části 5.4.1, kde bylo implementováno upřednostňované připojování v rámci celé sítě.



Obr. 26: Degree distribuce rostoucí síti

Vidíme, že rozdělení stupňů se zásadně změnilo a lze ho vcelku dobře, ač ne úplně přesně, aproximovat výše popisovaným mocninným zákonem. V každém případě tedy došlo ke zlepšení chování modelu směrem k reálným sociálním sítím.

Podívejme se v rámci degree distribuce ještě na 2D histogramy závislosti kvality agenta na jeho stupních, které jsou znázorněny na Obr. 27.



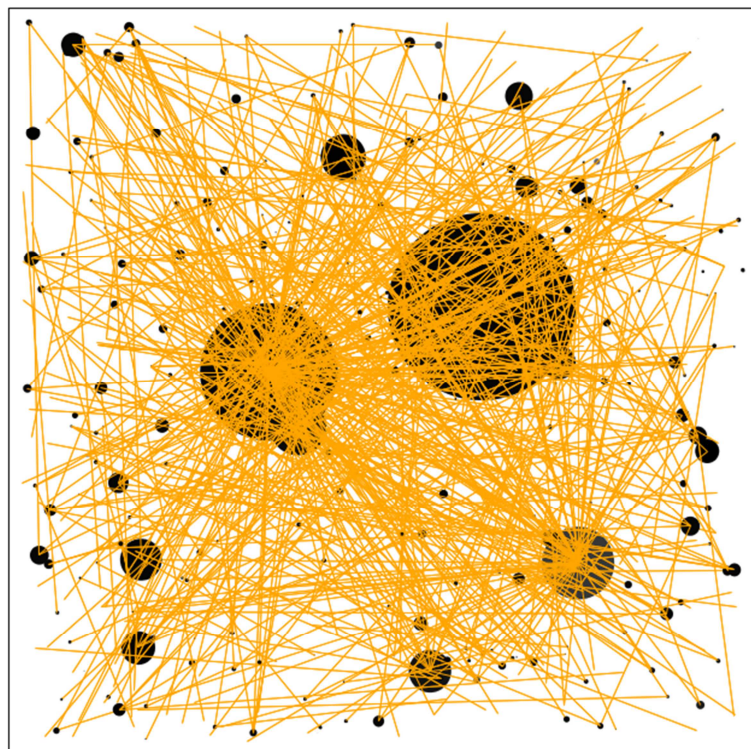
Obr. 27: 2D histogramy závislosti kvality agenta na počtu jeho kontaktů

Na osách x je uvedena normovaná hodnota odchozích a příchozích vazeb (*out/in-degree*) a na osách y potom kvalita agenta. Z histogramů lze usoudit, že čím více vazeb, bez ohledu na jejich orientaci, agent má, tím jsou jeho rozhodnutí kvalitnější. Je zřejmé, že čím více je partnerů (*outdegree*), kterých se může agent dotazovat, tím větší je šance získat kvalitní řešení. Naproti tomu, čím kvalitnější řešení sám agent poskytuje, tím více k sobě přitahuje ostatní agenty (*indegree*).

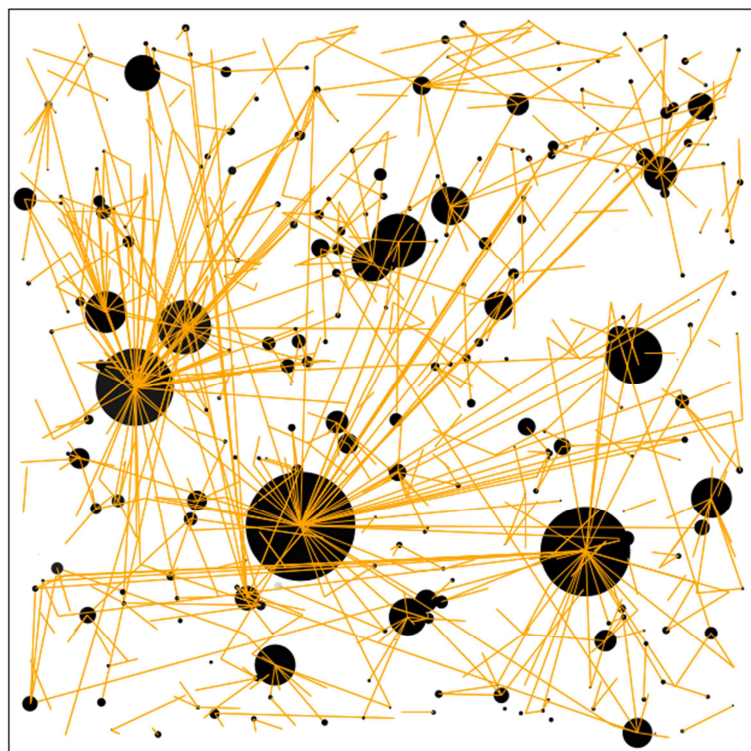
V dalším experimentu jsme se zaměřili na aplikaci upřednostňovaného připojování pouze na nejbližší okolí agenta, která se asi nejvíce projevila na struktuře sítě. Podotkneme, že v předchozím experimentu jsme na počátku simulace pracovali se sítí o 100 agentech, nyní je na počátku simulace v síti pouze jeden jediný agent, to znamená v kroku $t=1000$ obsahuje populace rovných 1000 agentů. Na Obr. 28 můžeme vidět 2D vizualizaci sítě s upřednostňovaným připojováním aplikovaným na celou síť.

Vidíme hustou strukturu s několika málo agenty, na které je napojena velká část sítě. Podobnou strukturu jsme mohli vidět již na Obr. 18. To odpovídá systému upřednostňovaného připojování. Naproti tomu se můžeme podívat na Obr. 29 znázorňující strukturu taktéž rostoucí sítě, ovšem již s lokálně aplikovaným upřednostňovaným připojováním.

Rozdíl je patrný na první pohled. V různých místech sítě začala vznikat menší centra, na která je napojeno větší množství uzlů, ale žádné není tak extrémní, aby pohlcovalo většinu sítě, jako tomu bylo v předchozím obrázku. To je logické, protože nově příchozí uzly se napojují pouze na jim nejbližší uzly a s největší pravděpodobností se napojí na takovéto jim nejbližší centrum, což celý obrázek jasně vysvětluje. Obr. 29 je také příkladem bezškálové struktury sítě



Obr. 28: Struktura rostoucí sítě s upřednostňovaným připojováním



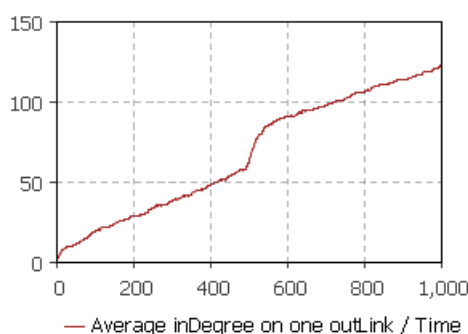
Obr. 29: Struktura rostoucí sítě s upřednostňovaným připojováním aplikovaným lokálně

Posledním experimentem zaměřeným na problematiku rostoucích sítí je eliminace mechanismu upřednostňovaného připojování známého ze zkoumání reálných sociálních sítí jakožto zdroje mocinného rozdělení stupňů a jeho nahrazení prostým napojením na m nejbližších jedinců. Předložený model tedy umožňuje zkoumat dynamiku tohoto mechanismu.

Jak je uvedeno výše, agent je po vstupu do sítě připojen k m svým nejbližším sousedům, v našem případě pro $m=5$, kde hodnoty stupňů jednotlivých uzlů nehrají žádnou roli. Efekt preferenčního připojení se zde tedy neprojevuje ihned v okamžiku příchodu agenta do sítě, ale v dynamice vývoje jeho vazeb, které agent upravuje k dosažení maximálního zisku z členství v síti. Pro demonstraci je na Obr. 30 uveden graf zachycující časový vývoj parametru d_{norm} , definovaného vzorcem (15).

$$d_{norm} = \frac{1}{N} \sum_{k=1}^N \frac{1}{P_k} \sum_{j=1}^{P_k} d_{ij} \quad (15)$$

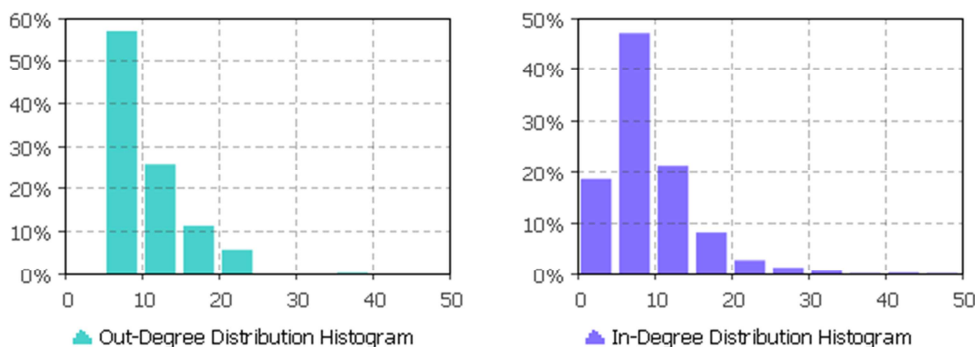
V něm je N celkový počet agentů v síti, P_k velikost seznamu partnerů agenta k a d_{ij} *indegree* partnera j ze seznamu partnerů P_k . Tento parametr vyjadřuje průměrnou kvalitu každé vazby v síti, resp. kvalitu partnera, ke kterému tato vazba směřuje. V případě potvrzení pravidla upřednostňovaného připojení bude tato hodnota v čase růst, což bylo testováno na výše definované rostoucí síti s výsledkem na Obr. 30.



Obr. 30: Průměrná kvalita odchozích vazeb v síti

Z grafu je zřejmé, že v průběhu simulace agenti sítě neustále zvyšovali kvalitu svých vazeb v tom smyslu, že je směřovali k partnerům s nejvyšším stupněm *indegree*. Právě v tomto procesu je možné pozorovat realizaci preferenčního připojování, které bylo

v teoretické části popsáno až ve svém finálním stavu. V grafu je patrné zvlnění sledované hodnoty. To ukazuje na skokové změny kvality vazeb, které nebyly vyvolány změnami nastavení modelu a lze je interpretovat jako nalezení významných (s vysokým *indegree*) jedinců v síti, na které se vzápětí orientovala celá řada dalších agentů svými vazbami. Podobné náhlé skokové změny jsme mohli pozorovat v průměru odchozích vazeb již na Obr. 24. Pojdme se ještě podívat na rozdělení hodnot stupňů v takto postaveném modelu. Výsledek můžeme vidět na Obr. 31.

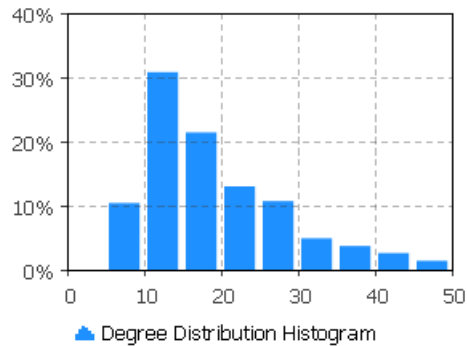


Obr. 31: Degree distribuce sítě s prostým napojováním na m nejbližších

Vidíme, že rozložení stupňů silně připomíná hyperbolický průběh mocninného zákona, ač tedy v pravém grafu se na začátku vyskytuje určitá odchylka. Zamysleme se ale ještě nad jednou věcí. Z teoretické části víme, že podle preferenčního připojování se uzly s největší pravděpodobností napojují na ty s nejvyšším počtem vazeb, ovšem již není specifikována jejich orientovanost, to znamená bez ohledu na jejich směr. I my jsme v implementaci preferenčního připojování pravděpodobnost založili na součtu odchozích a příchozích vazeb.

Podívejme se tedy ještě na Obr. 32, který znázorňuje degree distribuci bez ohledu na orientovanost vazeb v síti. Jde vlastně pouze o součet histogramů z Obr. 31.

Opět vidíme, že histogram svým tvarem velmi nápadně připomíná hyperbolu, ač tedy opět s malou nepřesností na začátku. Tato nepřesnost může být vysvětlena nastavením parametru m na hodnotu 5, čímž všichni agenti získají ihned po příchodu do sítě právě 5 vazeb, což může vyústit ve fakt, že agentů disponujících malým množstvím vazeb je v síti nižší počet.

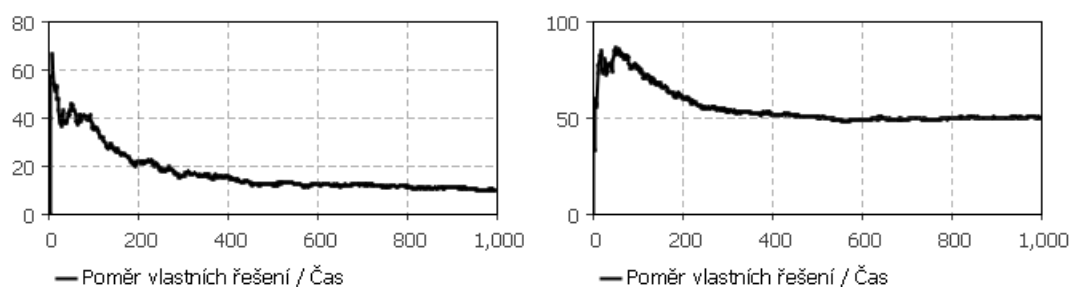


Obr. 32: Degree distribuce bez ohledu na orientovanost vazeb

V každém případě jsme se rozdělením hodnot stupňů v síti přiblížili mocninnému zákonu, a to i bez preferenčního připojování. To je důkaz toho, že samo upřednostňované připojování zohledňující stupně jednotlivých uzlů při příchodu dalších nemusí být jediným mechanismem vedoucím k rozdělení stupňů podle mocninného zákona. Jak jsme již řekli, ona preference může záviset na mnoha faktorech a vysoký stupeň některých uzlů zde není příčinou připojení, ale naopak důsledkem toho, že ony faktory směřují většinu sítě právě k těmto uzlům.

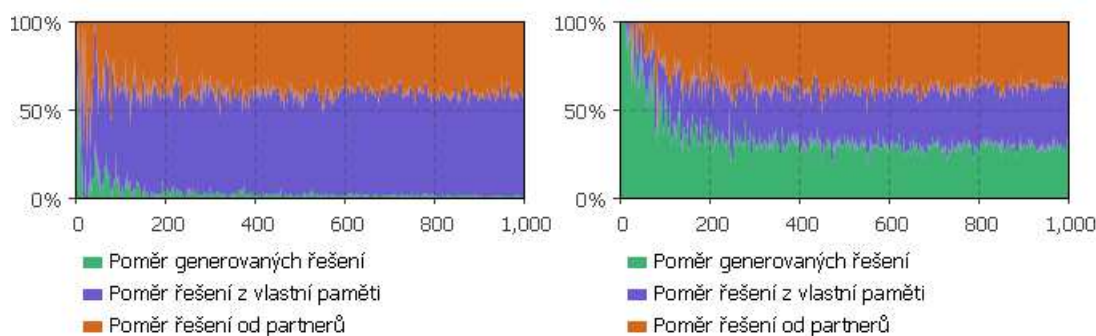
6.5 Flexibilní počet situací

Nyní se v rámci našeho experimentování podíváme na zavedení proměnného počtu situací. Ve všech předchozích experimentech jsme pracovali pouze s jednou situací vyžadující řešení. Následující experiment byl zaměřen na zkoumání vlivu úpravy modelu v oblasti řešení více situací současně. Model byl v tomto případě spuštěn ne s jednou, ale s dvaceti situacemi, další nastavení zůstala stejná. Na Obr. 33 je vidět porovnání poměru vlastních řešení v paměti pro 1 a 20 současně řešených situací. Na ose x je počet kroků simulace.



Obr. 33: Vývoj poměru vlastních řešení pro jednu a dvacet současně řešených situací

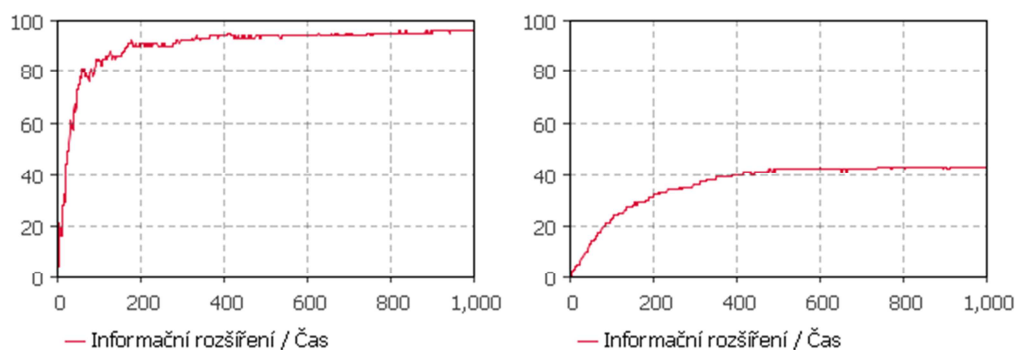
Levý graf vlastně odpovídá pravému grafu na Obr. 16. Můžeme tedy mimochodem podotknout, že již pracujeme s upraveným rozhodováním agentů. Vliv vyššího počtu řešených situací ukazuje pravý graf, v němž se poměr vlastních řešení v průběhu simulace ustálil kolem 50 procent. To znamená, průměrně jsou agenti autory celé poloviny všech řešení ve své paměti. To je logické, protože v síti se nyní řeší mnohem více situací a pro některé, respektive pro všechny, je k dispozici nedostatek řešení, a tudíž agentům nezbývá nic jiného než generovat svá vlastní. Můžeme si též všimnout, že graf po ustálení již nemá klesající charakter, který jsme vysvětlovali již v části 6.1. To taktéž souvisí s nedostatkem řešení pro tolik situací a je to také znak toho, že při takovémto množství řešených situací již nedochází k tak hladkému rozšíření znalostí. Je to zřejmě i z toho důvodu, že nyní se v síti silně projevuje také vliv zapomínání agentů, to znamená, než se stihnou některá řešení rozšířit po síti, jejich držitelé je zapomenou a všichni se tím pádem nevyhnou nutnosti generovat stále nová a nová řešení. Toho jest důkazem i Obr. 34 znázorňující grafy, které již také dobře známe z části 6.1.



Obr. 34: Poměry zdrojů řešení pro jednu a dvacet současně řešených situací

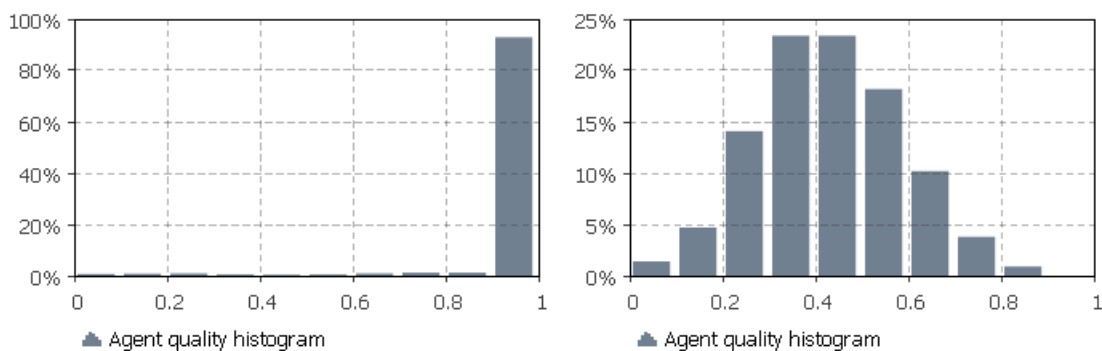
Levý graf opět, stejně jako v předchozím případě, odpovídá pravému grafu na Obr. 17. Z pravého grafu je zřejmé, že větší počet situací způsobil zejména na začátku větší, až 100 %, potřebu generovat vlastní řešení, adaptace sítě na složitější úlohu trvala déle a i v ustáleném stavu byl počet generovaných řešení stále mnohem vyšší. To odpovídá realitě – v případě většího počtu různých podnětů trvá optimalizace seznamu partnerů déle a celková kvalita řešení je nižší (agenti negenerují stejně kvalitní řešení pro všechny situace, počet vazeb je však omezen).

Zkoumán byl i vliv více situací na celkovou průměrnou kvalitu řešení poskytovaných agenty v celé síti. Na Obr. 35 jsou grafy vývoje kvality sítě jako celku pro jednu a dvacet situací.



Obr. 35: Kvalita sítě v čase pro jednu a dvacet současně řešených situací

Vidíme, že v případě řešení jedné situace se průměrná kvalita sítě vyšplhala velmi vysoko, a to až na 96 %. V případě řešení dvaceti situací naproti tomu síť dosáhla kvality pouze asi 43 %. To vše je v souladu s výše popisovanými jevy. K dispozici máme ještě histogramy rozložení kvality agentů na Obr. 36, který koreluje s Obr. 35.

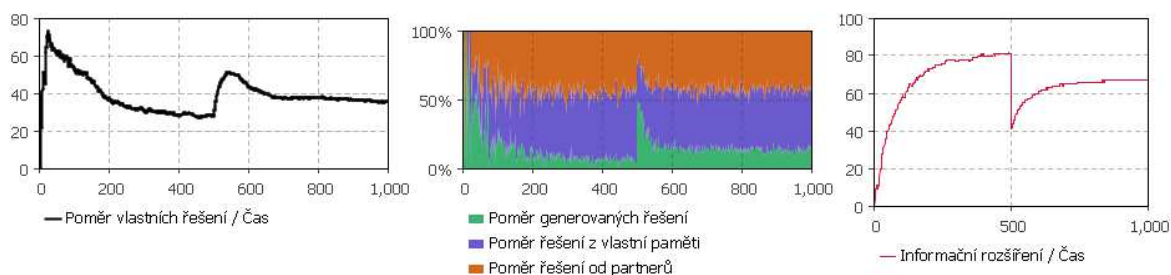


Obr. 36: Histogram kvality agentů pro jednu a dvacet současně řešených situací

Je zřejmé, že způsob nastavení ratingu pro agenta jako celek postačoval pro kvalitní nastavení sítě na řešení jedné situace, ne však dvaceti. Jako jedno z možných řešení se jeví úprava ratingového systému směrem k individuálnímu ratingu na agenta a situaci. Zde musíme ale zdůraznit, že nerozšíření veškerých znalostí ke všem agentům v síti nemůže být uvažováno jako znak nekvality simulace a opět, jako již jednou, musíme upozornit na to, že

naším cílem je co nejlépe simulovat reálný svět, ve kterém zkrátka nemohou všichni vědět úplně všechno.

Do modelu byla ještě implementována událost, která může v libovolném kroku simulace zvýšit počet řešených situací. Uvedme si tedy pro zajímavost ještě jeden experiment. Model byl spuštěn s pěti situacemi a přesně v polovině simulace, to znamená, v kroku 500 byl počet řešených situací náhle zvýšen o dalších pět, čili na deset. Na Obr. 37 je znázorněn výsledek.



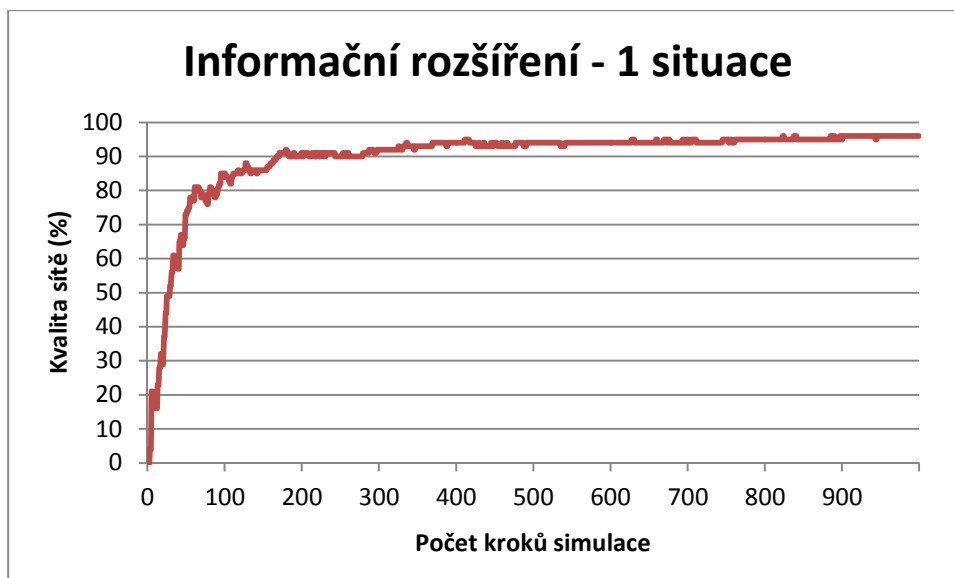
Obr. 37: Výsledek náhlé změny počtu řešených situací

Průběh grafů snad již není potřeba sáhodlouze vysvětlovat. Náhlé zdvojnásobení počtu situací vyvolalo potřebu generovat pro ně řešení, čímž pádem se začal prudce navyšovat poměr vlastních řešení v paměti agentů a logicky o polovinu klesla celková kvalita sítě.

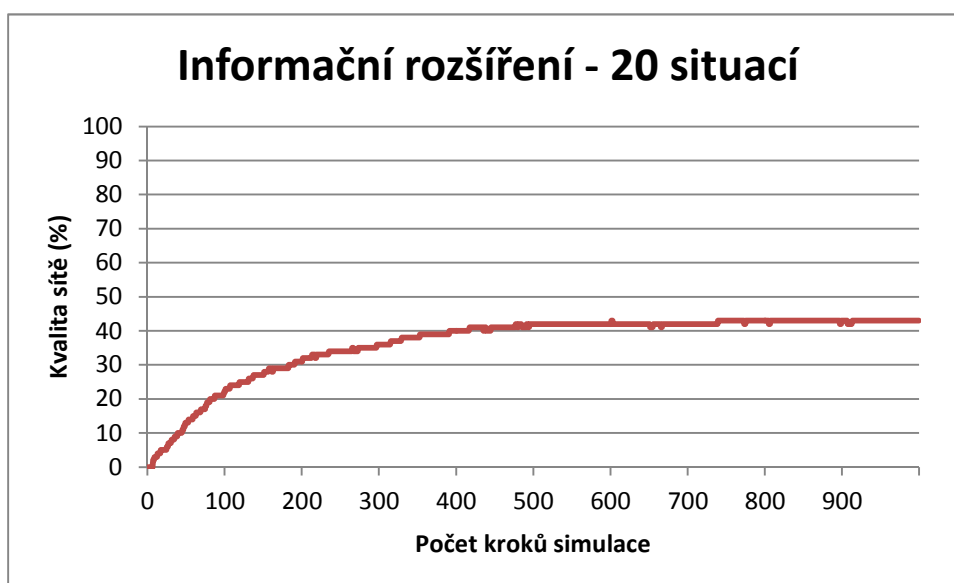
6.6 Analýza metodikou časových řad

Na závěr se pokusíme demonstrovat využití metodik, které jsme si uvedli v rámci teoretické části práce. Zamyslíme-li se, snímkování sítě provádí model vlastně sám, jelikož veškerá statistika je aktualizována v každém kroku simulace a výsledky jsou zanášeny do datových podkladů pro grafy. Zaznamenává-li model určitou charakteristiku v každém kroku, při délce simulace 1000 kroků nebude datový podklad obsahovat nic jiného než 1000 hodnot této charakteristiky seřazené chronologicky od prvního do posledního kroku, čili vlastně časovou řadu.

Pokusme se takto podívat na ukazatel průměrné kvality sítě. Využijeme k tomu program Excel, který do jisté míry umožňuje práci s časovými řadami. Pro zobrazení vývoje časové řady jsou typické spojnicové grafy, které můžeme vidět na Obr. 38 a Obr. 39 a které jsou shodné s Obr. 35.



Obr. 38: Graf kvality sítě v čase pro jednu situaci v programu Excel



Obr. 39: Graf kvality sítě v čase pro dvacet situací v programu Excel

6.6.1 Průměrný přírůstek znalostí

Průběh obou grafů je rostoucí, což je logické, protože informace se v průběhu času rozšiřují. Můžeme tedy změřit průměrný přírůstek kvality sítě po celou dobu simulace. Průměrný přírůstek vyjádříme podle vzorce (10). Výsledek zachycuje Tab. 2.

	Průměrný přírůstek kvality sítě
1 situace	0,096096096
20 situací	0,043043043

Tab. 2: Průměrný přírůstek kvality sítě pro jednu a dvacet situací

Vidíme, že průměrný přírůstek kvality je v případě řešení jedné situace větší než při řešení dvaceti situací, což je logické, protože znalosti o řešení jedné situace se rozšiřují snadněji než při současném řešení dvaceti situací, čemuž odpovídá i charakter grafů.

Prohlédneme-li si grafy pozorně, můžeme zjistit, že nárůst kvality sítě dosáhne v určitém momentě maxima, respektive stavu, kdy již nedochází k žádnému nebo jen nepatrnému nárůstu. Pokusme se nějakým způsobem identifikovat, kdy k tomuto, můžeme říci ustálení sítě, asi dojde. Napovědět nám může Tab. 3.

Kroky simulace	Průměrný přírůstek kvality sítě	
	1 situace	20 situací
0-100	0,858586	0,222222
100-200	0,060606	0,090909
200-300	0,010101	0,050505
300-400	0,020202	0,040404
400-500	0	0,020202
500-600	0	0
600-700	0,010101	0
700-800	0	0,010101
800-900	0	0
900-1000	0,010101	0

Tab. 3: Průměrný přírůstek kvality sítě za každých sto kroků simulace

Celá doba simulace byla rozdělena na 10 intervalů, tzn. po 100 krocích, a průměrný přírůstek kvality byl vypočítán pro každý tento interval zvlášť. Tím se dozvídáme zajímavé výsledky. Na první pohled je zřejmé, že k největšímu nárůstu kvality dochází během prvních 100 kroků simulace. Ve zbytku simulační doby již nikdy k tak velkému nárůstu nedochází a lze usoudit, že zvyšování znalostí se v průběhu simulace neustále snižuje.

Podívejme se pozorně na Tab. 3. Jak již bylo zmíněno, v případě řešení jedné situace dosáhne kvalita sítě na konci simulace 96 %, ovšem již v kroku 400 dosahuje cca 94 %. V případě dvaceti řešených situací dosáhne kvalita sítě na konci 43 %, ale i zde již v polovině simulační doby, čili v kroku 500, dosahuje 42 %. V rámci těchto úvah bychom mohli říci, že při řešení jedné situace dojde k ustálení zhruba v kroku 400 a při řešení dvaceti v kroku 500 nebo obecně, že v druhé polovině simulace již téměř k žádnému navyšování znalostí nedochází. Usoudit lze i to, že v případě řešení dvaceti situací je navyšování znalostí pozvolnější.

6.6.2 Dekompozice vývoje kvality sítě

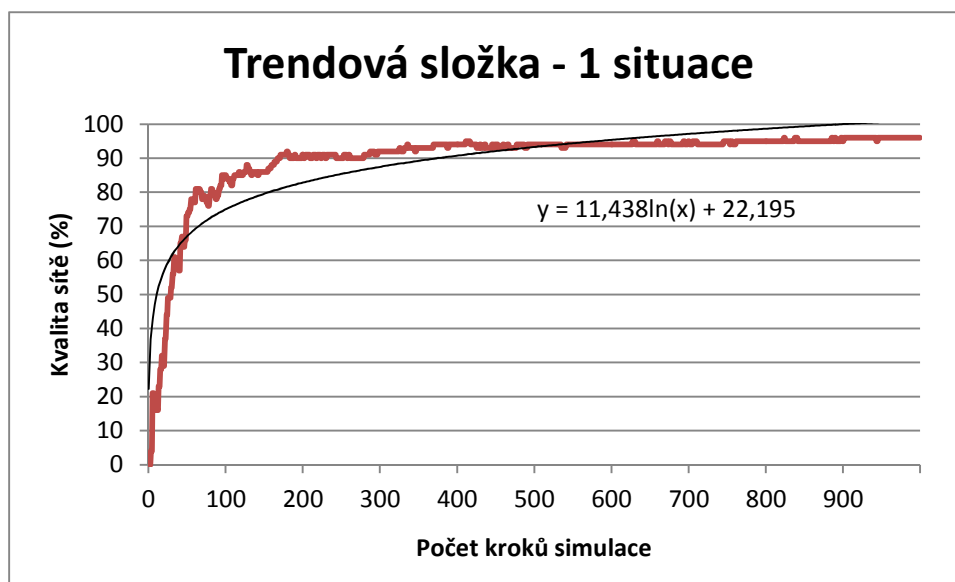
Předpokládejme aditivní model dekompozice a zkusme časové řady na Obr. 38 a Obr. 39 rozložit na jednotlivé složky podle vzorce (8). V případě našeho simulačního modelu nepředpokládáme žádnou sezónní ani cyklickou složku, tudíž je vypustíme a časové řady rozložíme podle vzorce (16). Předpokládáme tedy pouze trendovou a náhodnou složku.

$$y_t = T_t + \varepsilon_t \quad (16)$$

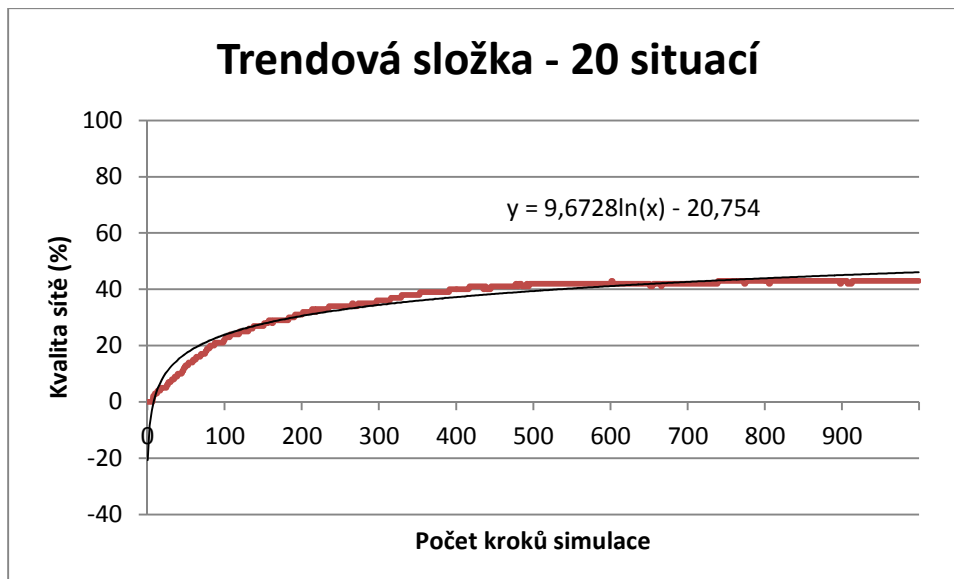
Nejvíce nás bude zajímat trendová složka, která nám odhalí obecnou tendenci vývoje kvality sítě. Náhodná složka zbyde po vyloučení trendové složky. Trend určíme analytickým způsobem, kterým získáme popis matematické funkce. Opět využijeme programu Excel, který toto umožňuje. Trend nejlépe popisuje logaritmická funkce, která má vzorec (17).

$$T_t = a_0 \ln t + a_1 \quad (17)$$

Výsledek můžeme vidět na Obr. 40 a Obr. 41, kde jsou zároveň napsány konkrétní rovnice trendových funkcí.

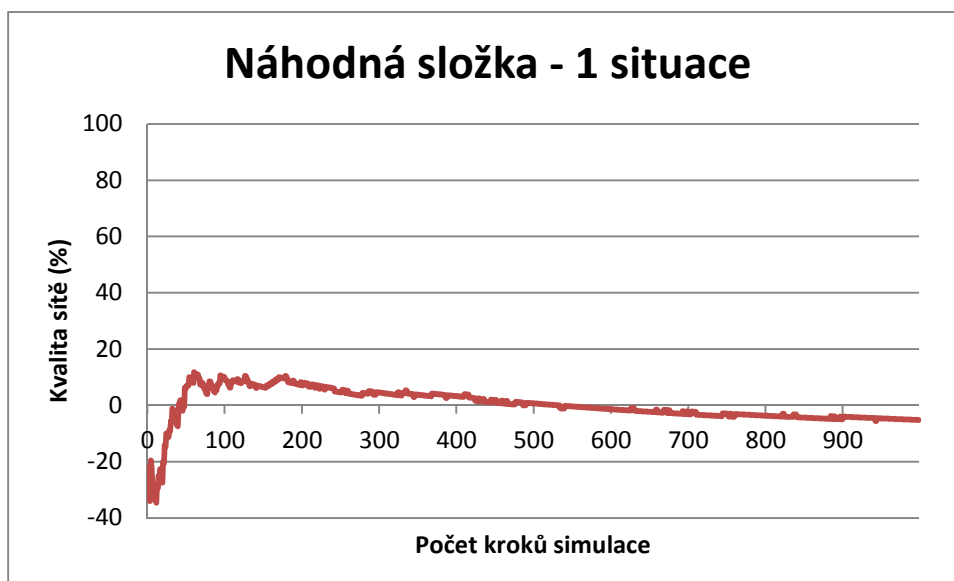


Obr. 40: Graf trendu vývoje kvality sítě pro jednu situaci

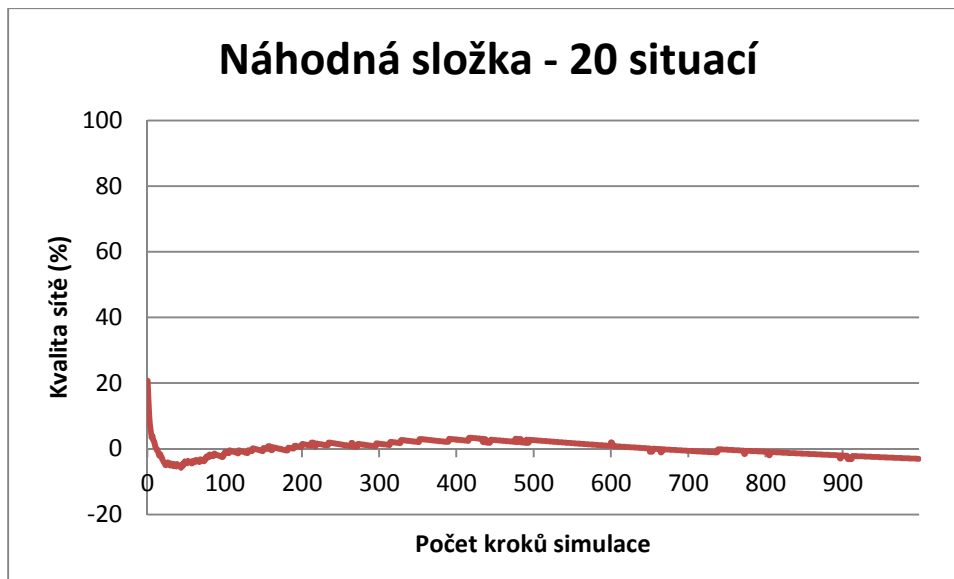


Obr. 41: Graf trendu vývoje kvality sítě pro dvacet situací

Na Obr. 42 a Obr. 43 si ještě můžeme prohlédnout náhodné složky, které vzniknou odečtením trendových složek.



Obr. 42: Graf náhodné složky kvality sítě pro jednu situaci



Obr. 43: Graf náhodné složky kvality sítě pro dvacet situací

6.6.3 Směrnice tečny jako metrika ustálení sítě

V předchozí části jsme se snažili identifikovat, kdy dojde k ustálení kvality sítě. Nyní máme k dispozici matematický popis trendu, který se pokusíme k této identifikaci využít. Logaritmické funkce, které můžeme vidět na Obr. 40 a Obr. 41, jsou rostoucí, přičemž změna hodnoty funkce se s rostoucím x neustále snižuje, tzn., sestrojíme-li tečnu k tomuto grafu podle vzorce (13) v určitém bodě x_0 a tento bod budeme posouvat po grafu směrem doprava, bude se sklon tečny snižovat a v momentě, kdy dosáhne určité hodnoty, můžeme předpokládat, že je síť ustálená, tudíž směrnici tečny můžeme využít jako metriku ustálení.

Otázkou je, jaká má být hodnota sklonu. Podle výše prováděného výpočtu průměrného přírůstku kvality jsme usoudili, že k ustálení dojde zhruba v polovině simulační doby. Můžeme tedy zjistit směrnici tečny v tomto momentě. Výsledek zachycuje Tab. 4, kde najdeme i další hodnoty směrnic v průběhu simulace.

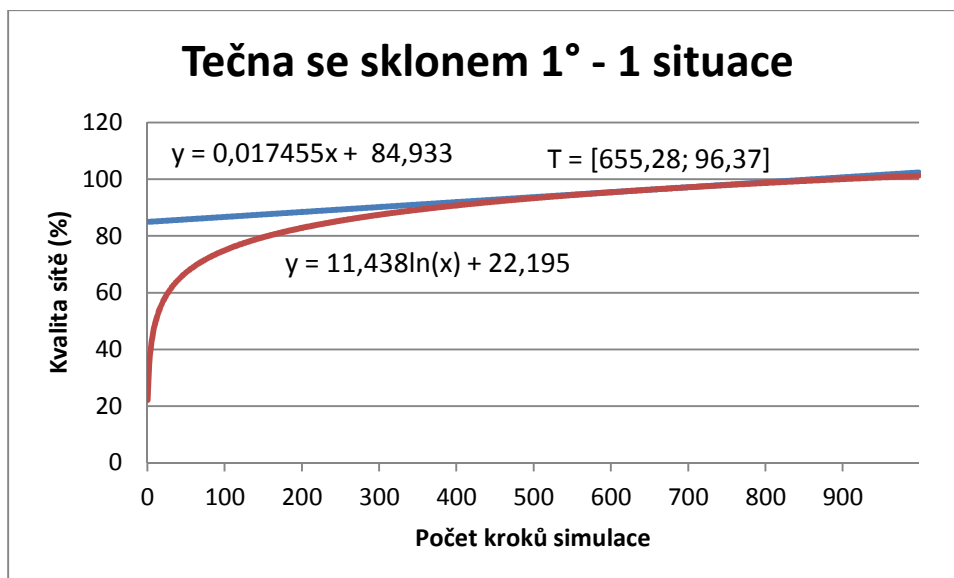
Krok simulace	1 situace		20 situací	
	Směrnice ($\tan \alpha$)	Úhel (α)	Směrnice ($\tan \alpha$)	Úhel (α)
100	0,11438	6,5251°	0,096728	5,5249°
200	0,05719	3,2732°	0,048364	2,7689°
300	0,038126667	2,1834°	0,032242667	1,8467°
400	0,028595	1,6379°	0,024182	1,3853°
500	0,022876	1,3105°	0,0193456	1,1083°
600	0,019063333	1,0921°	0,016121333	0,9236°
700	0,01634	0,9361°	0,013818286	0,7917°
800	0,0142975	0,8191°	0,012091	0,6927°
900	0,012708889	0,7281°	0,010747556	0,6158°
1000	0,011438	0,6553°	0,0096728	0,5542°

Tab. 4: Sklon tečny trendové funkce pro jednu a dvacet situací v průběhu simulace

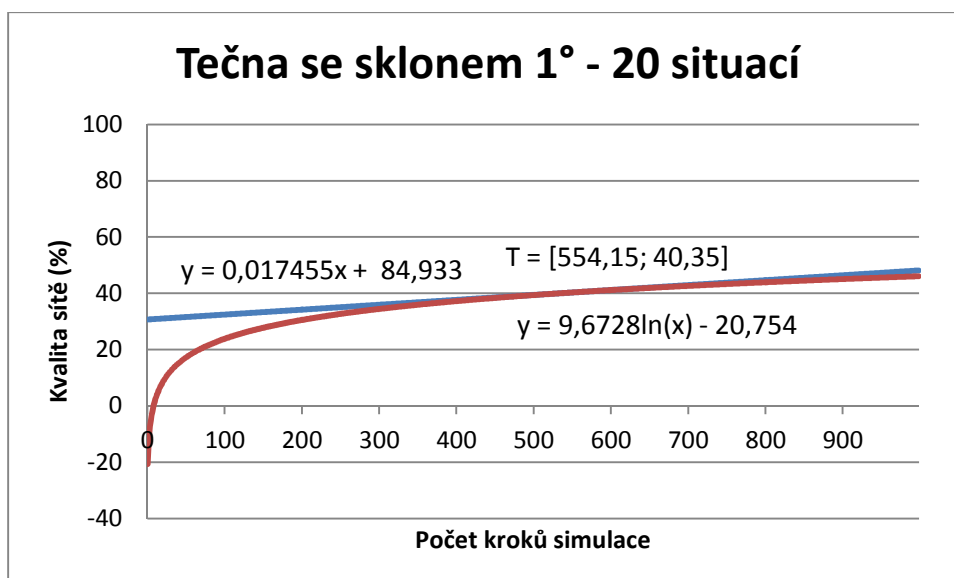
Směrnici tečny lze vypočítat, jak můžeme vidět v Tab. 4, také jako *tangens* úhlu, který tečna svírá s osou x . Podobně jako v případě průměrného přírůstku je změna respektive snížení sklonu největší začátkem simulace. Vidíme, že v polovině simulace svírá tečna s osou x v obou případech něco málo přes jeden stupeň a je s ní tedy téměř rovnoběžná. V dalším průběhu se sklon snižuje již velmi málo. Mohli bychom tedy stanovit, že k ustálení kvality sítě dojde v momentě, kdy tečna dosáhne sklonu jednoho stupně. Použijeme k tomu opět vzorec (13), s tím rozdílem, že nyní nehledáme směrnici, tu známe, ale bod dotyku. Výsledek je v Tab. 5 a na Obr. 44 a Obr. 45, kde jsou rovněž napsány vzorce trendových funkcí, jejich tečen a také body dotyku T .

Úhel (α)	Krok simulace	
	1 situace	20 situací
1°	655,28	554,15

Tab. 5: Moment dosažení zadaného sklonu tečny



Obr. 44: Graf momentu dosažení zadaného sklonu tečny pro jednu situaci



Obr. 45: Graf momentu dosažení zadaného sklonu tečny pro dvacet situací

Z bodu dotyku T můžeme jednoduše zjistit, kdy (souřadnice x) a k jakému rozšíření znalostí (souřadnice y) došlo při ustálení sítě. Někdo by mohl namítnout, že podle výsledků na Obr. 44 a Obr. 45 dojde k ustálení sítě dříve při řešení dvaceti situací než při řešení jedné. Musíme si ale uvědomit, že měření bylo prováděno na obecném trendu vývoje časových řad, znázorněných na Obr. 40 a Obr. 41, kdy z důvodu umožnění použití derivací musely být vypuštěny náhodné složky, znázorněné na Obr. 42 a Obr. 43, které bohužel nelze popsat žádnou funkcí času.

7 Závěr

V naší práci *Dynamika sociálních sítí* jsme usilovali o naplnění několika cílů. V první řadě jsme si představili pojem sociální sítě, jeho vznik, význam a chápání v dnešním světě, a řekli jsme si něco o základních vlastnostech sociálních sítí a problematice jejich analýzy. S rozvojem informačních technologií, které nám umožňují zkoumání rozsáhlých sítí, se dostáváme k pojmu komplexní sítě, pod který spadají i sítě sociální. Tyto sítě mají mnoho společných vlastností, které je možné sledovat s využitím řady metrik, jejichž přehled byl vytvořen v rámci práce a z nichž nejvíce používané jsou rozdělení stupňů, shlukování a průměrná nejkratší cesta v síti. Dále jsme si uvedli tři základní teoretické modely síťové struktury, které vznikly v průběhu vývoje síťové analýzy, a to model náhodného grafu, který je považován za jeden z nejstarších, dále model malého světa a nakonec modely *scale-free* sítí, které již na rozdíl od předchozích dvou zahrnují vývoj síťové struktury v čase neboli dynamiku.

V další části práce jsme se zaměřili na aktuální informační zdroje zabývající se dynamikou sociálních sítí a nastínili si směry, kterými se tato problematika ubírá.

Další část práce byla věnována metodikám modelování a zkoumání dynamiky. Uvedli jsme si dva možné způsoby modelování. Prvním z nich byla systémová dynamika, která se obvykle používá při globálním pohledu na dynamiku systému, kdy jsou jednotlivé prvky v modelu zastoupeny pouze podle jejich množství. Druhým způsobem bylo agentové modelování, které lze charakterizovat jako individuálně orientované, kdy se globální chování projevuje jako interakce jednotlivých chování aktivních subjektů, tzv. agentů. Co se týká zkoumání dynamiky sociálních sítí, uvedli jsme si tzv. snímkování sítě, které spočívá v diskretizaci času a zachycení vývoje sítě do jednotlivých snímků, které jsou poté porovnávány, a metody analýzy časových řad, které právě tímto snímkováním mohou vznikat.

V další části práce, praktické, jsme se seznámili s již dříve navrženým simulačním modelem pro zkoumání dynamiky informačních toků v sociálních sítích. Podrobný popis modelu je možné nalézt na příloženém CD. Tento model byl s přihlédnutím na nabyté teoretické poznatky a logické úvahy autora v rámci práce dále rozvinut a zdokonalen.

Bylo vylepšeno rozhodování jedince v oblasti získávání znalostí ze sítě, dále byla upravena inicializace ochoty jedinců odpovídat na jim kladené dotazy, s čímž souvisí i implementace klesající důvěry v jedince, kteří ochotni odpovídat nejsou. Dále byl do modelu implementován fakt, že jedinci si pamatují, kým byli dotazováni a jsou schopni tento kontakt

následné využití, čímž byl mimochodem také vyřešen problém jednosměrnosti vazeb, který se vyskytl po implementaci rostoucí sítě, která představuje další modifikaci modelu. Rostoucí síť byla implementována prostřednictvím upřednostňovaného připojování, díky kterému se rozdělení stupňů v síti přiblížilo mocninnému zákonu známému z reálných sociálních sítí. Poslední úpravou bylo zavedení flexibilního počtu řešených situací v síti. V rámci vyššího počtu situací nyní jedinci nedisponují pouze jedním parametrem kvality, ale kvalita či inteligence jedinců je nastavována zvlášť pro každý typ situace.

V poslední části práce byla provedena řada experimentů, které demonstrují jednak vliv všech provedených úprav na výstupy modelu, ale také možnost jeho užití pro zkoumání dynamiky sociálních sítí. Všechny experimenty jsou doprovázeny diskuzí vysvětlující dynamické jevy, které se v průběhu simulací vyskytovaly.

Závěrem lze říci, že upravený model nyní podstatně lépe simuluje chování lidí v rámci komplexních sítí, ovšem nutno podotknout, že jistě ne dokonale, protože čelíme nelehkému úkolu, a tím je matematický popis dění v reálném světě. Složitost reálného světa je tak velká, že nalézt jeho dokonalý popis snad ani není možné. Definice modelu ovšem zní: „Model je zjednodušením reality.“ A cílem každého kvalitního modelu by měla být minimalizace tohoto zjednodušení, což bylo i cílem naší diplomové práce, a tento cíl byl splněn.

Seznam použité literatury

- [1] Valerio Arnaboldi, Marco Conti, Andrea Passarella, and Robin Dunbar. 2013. Dynamics of personal social relationships in online social networks: a study on twitter. In *Proceedings of the first ACM conference on Online social networks (COSN '13)*. ACM, New York, NY, USA, 15-26. DOI=10.1145/2512938.2512949 <http://doi.acm.org/10.1145/2512938.2512949>
- [2] GÁLIK, M. *Models and Methods in Complex Networks*. České Budějovice 2012. Bachelor's thesis. South-Bohemian University, České Budějovice. Faculty of Education. ICT Department. Supervisor: Ing. Ladislav Beránek, CSc.
- [3] Klimeš, R. 2012. *Využití sociálních sítí v organizaci*. [The Use of Social Network in Organization. Bc. Thesis, in Czech.] – 49 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.
- [4] Michael Steurer and Christoph Trattner. 2013. Acquaintance or partner?: predicting partnership in online and location-based social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*. ACM, New York, NY, USA, 372-379. DOI=10.1145/2492517.2492562 <http://doi.acm.org/10.1145/2492517.2492562>
- [5] Materiály z přednášek předmětu UAI/689 – *Algoritmy II* na PřF JU v Českých Budějovicích, 2010; přednášející doc. Ing. Ladislav Beránek, CSc., MBA, Katedra informatiky, PF JU v Českých Budějovicích.
- [6] Česko-Slovenská filmová databáze © 2001-2015, *Šest stupňů odloučení*, [Online]. 1993 [cit. 2015-12-09]. Dostupné z: <http://www.csfd.cz/film/14780-sest-stupnu-odlouceni/prehled/>
- [7] Miray Kas, Kathleen M. Carley, and L. Richard Carley. 2013. Incremental closeness centrality for dynamically changing social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '13)*. ACM, New York, NY, USA, 1250-1258. DOI=10.1145/2492517.2500270 <http://doi.acm.org/10.1145/2492517.2500270>
- [8] Wasserman, S., Faust, K. *Social network analysis : methods and applications*. Cambridge University Press, 2007. ISBN 978-0521-38707-1
- [9] Xiaohan Zhao, Alessandra Sala, Christo Wilson, Xiao Wang, Sabrina Gaito, Haitao Zheng, and Ben Y. Zhao. 2012. Multi-scale dynamics in a massive online social network. In *Proceedings of the 2012 ACM conference on Internet measurement conference (IMC '12)*. ACM, New York, NY, USA, 171-184. DOI=10.1145/2398776.2398795 <http://doi.acm.org/10.1145/2398776.2398795>

- [10] Luca Allodi, Luca Chiodi, and Marco Cremonini. 2011. The asymmetric diffusion of trust between communities: simulations in dynamic social networks. In *Proceedings of the Winter Simulation Conference (WSC '11)*, S. Jain, R. Creasey, J. Himmelspach, K. P. White, and M. C. Fu (Eds.). Winter Simulation Conference 3146-3157.
- [11] Pelánek, R., *Modelování komplexních sítí*, [Online]. [cit. 2015-05-23]. Dostupné z: <http://www.fi.muni.cz/~xpelane/IV109/slidy/networks.pdf>
- [12] Newman, M., Barabási, A. & Watts, D. 2006. *The structure and dynamics of networks*; Mark Newman, Albert-László Barabási, Duncan J. Watts, editors. Princeton: Princeton University Press.
- [13] Tanya Y. Berger-Wolf and Jared Saia. 2006. A framework for analysis of dynamic social networks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '06)*. ACM, New York, NY, USA, 523-528. DOI=10.1145/1150402.1150462 <http://doi.acm.org/10.1145/1150402.1150462>
- [14] Erheng Zhong, Wei Fan, Yin Zhu, and Qiang Yang. 2013. Modeling the dynamics of composite social networks. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13)*, Inderjit S. Dhillon, Yehuda Koren, Rayid Ghani, Ted E. Senator, Paul Bradley, Rajesh Parekh, Jingrui He, Robert L. Grossman, and Ramasamy Uthurusamy (Eds.). ACM, New York, NY, USA, 937-945. DOI=10.1145/2487575.2487652 <http://doi.acm.org/10.1145/2487575.2487652>
- [15] *PSLC Datashop* [Online]. 2015, 2015-10-02 [cit. 2015-12-09]. Dostupné z: <https://pslclatashop.web.cmu.edu/>
- [16] AnyLogic Company, 1992-2015. AnyLogic. [Online] Available at: <http://www.anylogic.com>
- [17] Hančlová, J., Tvrđý, L. *Úvod do analýzy časových řad*, [Online]. 2003 [cit. 2015-11-30]. Dostupné z: http://gis.vsb.cz/pan-old/Skoleni_Texty/TextySkoleni/AnalyzaCasRad.pdf
- [18] Jelínek, J. 2011. Modelování informačních toků v sociálních sítích. Znalosti 2011 - Czech and Slovak Knowledge Technology Conference, Stará Lesná, SR, 31. 1. 2011 - 2. 2. 2011. In *CEUR Workshop Proceedings*, Vol-802, <http://ceur-ws.org/Vol-802/> (online), ISBN: 978-80-248-2369-0

Seznam obrázků

Obr. 1: Sociální síť jako graf propojených uzlů	8
Obr. 2: Ukázka výpočtu koeficientu shlukování	11
Obr. 3: Distribuce stupňů podle mocinného zákona (Power law).....	12
Obr. 4: Graf s barevně odlišenými komunitami odpovídající matici sousednosti z Tab. 1....	16
Obr. 5: Tři náhodné grafy pro různá p a různá m	17
Obr. 6: Model malého světa vzniká kombinací pravidelného a náhodného grafu	19
Obr. 7: Příklad e-mailu a jeho zpracování	21
Obr. 8: Dva dynamické grafy, které se mapují na stejný statický graf.....	22
Obr. 9: Model ego sítě	24
Obr. 10: Modelování populace lišek a zajíců metodou systémové dynamiky	28
Obr. 11: Model v softwarovém prostředí AnyLogic 6.9.0	35
Obr. 12: Main – hlavní část modelu	35
Obr. 13: Agent Individual.....	36
Obr. 14: Rozšíření modelu o další funkcionalitu.....	37
Obr. 15: Rozšíření agenta Individual.....	38
Obr. 16: Vývoj poměru vlastních řešení před a po úpravě způsobu rozhodování agentů.....	45
Obr. 17: Poměry zdrojů řešení před a po úpravě způsobu rozhodování agentů.....	46
Obr. 18: Struktura sítě dle původního modelu	48
Obr. 19: Struktura sítě dle inovovaného modelu	48
Obr. 20: 2D Histogram závislosti ochoty agenta odpovídat na jeho oblibě	49
Obr. 21: Vývoj velikosti seznamů partnerů před úpravou modelu	50
Obr. 22: Struktura statické sítě původního modelu	51
Obr. 23: Problém orientovanosti vazeb	52
Obr. 24: Vývoj velikosti seznamů partnerů po úpravě modelu	52
Obr. 25: Degree distribuce na statické síti.....	54
Obr. 26: Degree distribuce rostoucí sítě	54
Obr. 27: 2D histogramy závislosti kvality agenta na počtu jeho kontaktů.....	55
Obr. 28: Struktura rostoucí sítě s upřednostňovaným připojováním.....	56
Obr. 29: Struktura rostoucí sítě s upřednostňovaným připojováním aplikovaným lokálně ...	56
Obr. 30: Průměrná kvalita odchozích vazeb v síti.....	57
Obr. 31: Degree distribuce sítě s prostým napojováním na m nejbližších	58
Obr. 32: Degree distribuce bez ohledu na orientovanost vazeb	59

Obr. 33: Vývoj poměru vlastních řešení pro jednu a dvacet současně řešených situací.....	59
Obr. 34: Poměry zdrojů řešení pro jednu a dvacet současně řešených situací.....	60
Obr. 35: Kvalita sítě v čase pro jednu a dvacet současně řešených situací.....	61
Obr. 36: Histogram kvality agentů pro jednu a dvacet současně řešených situací.....	61
Obr. 37: Výsledek náhlé změny počtu řešených situací.....	62
Obr. 38: Graf kvality sítě v čase pro jednu situaci v programu Excel.....	63
Obr. 39: Graf kvality sítě v čase pro dvacet situací v programu Excel.....	63
Obr. 40: Graf trendu vývoje kvality sítě pro jednu situaci.....	65
Obr. 41: Graf trendu vývoje kvality sítě pro dvacet situací.....	66
Obr. 42: Graf náhodné složky kvality sítě pro jednu situaci.....	66
Obr. 43: Graf náhodné složky kvality sítě pro dvacet situací.....	67
Obr. 44: Graf momentu dosažení zadaného sklonu tečny pro jednu situaci.....	69
Obr. 45: Graf momentu dosažení zadaného sklonu tečny pro dvacet situací.....	69

Seznam tabulek

Tab. 1: Příklad matice sousednosti pro výpočet modularity.....	15
Tab. 2: Průměrný přírůstek kvality sítě pro jednu a dvacet situací.....	63
Tab. 3: Průměrný přírůstek kvality sítě za každých sto kroků simulace.....	64
Tab. 4: Sklon tečny trendové funkce pro jednu a dvacet situací v průběhu simulace.....	68
Tab. 5: Moment dosažení zadaného sklonu tečny.....	68

Příloha

Obsah přiloženého CD

Na přiloženém CD lze nalézt:

- text diplomové práce ve formátu pdf
- popis původního modelu ve formátu pdf
- projekt s upraveným modelem ve formátu alp